

Analyzing word: cycling	
Top 10 similar words (custom model):	
Similar Word	Similarity
recycling	0.8962320685386658
gambling	0.8421981930732727
sailing	0.8114925622940063
ling	0.8052074313163757
dueling	0.7980522513389587
mailing	0.7960039377212524
yeungling	0.7588538527488708
feeling	0.7400113344192505
deejaying	0.7093671560287476
flying	0.700059175491333
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njpjphb	-0.02863088995218277
ever	-0.06275083124637604
holy	-0.10470271110534668
make	-0.11629704385995865
burrito	-0.11713263392448425
restaurant	-0.11926502734422684
else	-0.12016291916370392

good	-0.12170646339654922
breakfast	-0.12177065759897232
lunch	-0.12249930948019028
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
cyling	0.7613970637321472
bicycling	0.7530921697616577
Cycling	0.7459410429000854
biking	0.7440891861915588
cycling.The	0.7203680872917175
non-cycling	0.7042672038078308
cyclists	0.6933119297027588
cyclist	0.6866253018379211
pro-cycling	0.6842337250709534
cycling.	0.6818845272064209
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Sharef	0.23892761766910553
Semaj	0.23661378026008606
Ashbel	0.21913579106330872
Sango	0.21866606175899506

Zaron	0.21659745275974274
Olevia	0.2099568396806717
IMMANUEL	0.20697055757045746
Sarofim	0.2050960212945938
Ayoka	0.20497317612171173
Uchi	0.20187833905220032
Analyzing word: medallion	
Top 10 similar words (custom model):	
Similar Word	Similarity
scallion	0.7722135782241821
dillion	0.737381100654602
million	0.7128214836120605
lion	0.6855007410049438
elliot	0.6848589181900024
cuon	0.679830014705658
kabab	0.6697196364402771
cordon	0.6659064292907715
asopada	0.6544989347457886
sazon	0.6526552438735962
Top 10 dissimilar words (custom model):	

Dissimilar Word	Dissimilarity
back	-0.10879651457071304
dining	-0.11398310959339142
table	-0.11647014319896698
njpjphb	-0.13174964487552643
line	-0.13956046104431152
food	-0.1450989842414856
vibe	-0.14842189848423004
pleasant	-0.15009286999702454
space	-0.15119367837905884
prepared	-0.15256334841251373
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
medallions	0.816379725933075
medalion	0.7106068134307861
Medallion	0.6476216316223145
Medallions	0.5871028304100037
medalions	0.5828197002410889
medaillon	0.5625349879264832
pendant	0.5425906181335449
brooch	0.5405716896057129
emblem	0.5369361042976379

bas-relief	0.5264031291007996
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Append	0.20697833597660065
www.engvid.com	0.19864238798618317
released.Sign	0.1914958357810974
StaffBruin	0.19041572511196136
FYour	0.18825961649417877
pcper.com	0.18572098016738892
3Introduction	0.1852504014968872
UsTips	0.18489420413970947
LiNKs	0.18434756994247437
hamishcampbell.com	0.18355560302734375
Analyzing word: peppercorn	
Top 10 similar words (custom model):	
Similar Word	Similarity
peppers	0.8314974308013916
pepper	0.8204990029335022
peppermill	0.8045063018798828
pepperoni	0.785993754863739

peppermint	0.7601097822189331
szechuan	0.7097110152244568
sichuan	0.6638644933700562
asiago	0.6633860468864441
jumbulya	0.6608521342277527
pepe	0.6233408451080322
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
floor	-0.08772927522659302
variety	-0.10516609996557236
register	-0.11070123314857483
bartender	-0.11301339417695999
back	-0.11401569098234177
head	-0.11576472967863083
stay	-0.12009832262992859
event	-0.1244058683514595
display	-0.1255609691143036
post	-0.12692077457904816
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
peppercorns	0.676833987236023

pepper	0.6110559105873108
black-pepper	0.5607118010520935
Peppercorn	0.5477286577224731
pepper.The	0.5449494123458862
truffle	0.5279898047447205
Peppercorns	0.5276195406913757
lemon-grass	0.5237173438072205
nutmeg	0.5189248323440552
coriander	0.512756884098053
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
MiamiHerald.com.	0.2230479121208191
Somos	0.1785961538553238
-96	0.17704512178897858
Britani	0.17667600512504578
Komaraju	0.17443682253360748
Jockusch	0.17372965812683105
1-800-273-8255	0.1690007746219635
Gergo	0.16848056018352509
Despierta	0.16838213801383972
Codreanu	0.168363556265831

Analyzing word: franklin	
Top 10 similar words (custom model):	
Similar Word	Similarity
franklinville	0.8067992925643921
frank	0.805891215801239
fran	0.7603361010551453
frankie	0.7531478404998779
fraud	0.6737109422683716
frap	0.6545419096946716
franzone	0.6162093281745911
francisco	0.6071328520774841
frapp	0.5475205779075623
frame	0.5463414192199707
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njpjphb	-0.021679116412997246
come	-0.02958614006638527
potato	-0.05297166481614113
ever	-0.08774127066135406
extra	-0.08960167318582535
sauce	-0.09396318346261978

rice	-0.0996098518371582
world	-0.1026921197772026
tender	-0.1041940450668335
huge	-0.10714737325906754
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
jefferson	0.6322324275970459
jackson	0.6287664771080017
bradley	0.6058233976364136
hancock	0.6024911999702454
adams	0.597802460193634
rutherford	0.593543291091919
moore	0.5931870341300964
lyons	0.5879991054534912
davison	0.5877336859703064
harris	0.5866327285766602
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Subchannel	0.23833559453487396
Full-Body	0.225263774394989
Droplet	0.22376224398612976

VHD	0.21627548336982727
90-Minute	0.2107868641614914
go.For	0.20609880983829498
Deployment	0.20229341089725494
readersAllTwitter	0.20149624347686768
Pre-Post	0.1974009871482849
Enterovirus	0.19728441536426544

Analyzing word: fabulous	
Top 10 similar words (custom model):	
Similar Word	Similarity
fantabulous	0.8092814087867737
ridonkulous	0.6148379445075989
meticulous	0.5602449178695679
lousy	0.522161066532135
humongous	0.513922929763794
ridiculous	0.4923698604106903
angus	0.4797841012477875
marvelous	0.4776155054569244
luscious	0.4581413865089416
continuous	0.4568721354007721

Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
boat	-0.05046964064240455
compare	-0.0509353168308734
wall	-0.0524965487420558
young	-0.0553504787385463
free	-0.0612751357257366
subway	-0.0631222501397132
community	-0.0653014481067657
clearance	-0.0695885643362999
please	-0.0701043307781219
ramp	-0.0723265707492828
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
fantastic	0.8131504654884338
gorgeous	0.8097357749938965
wonderful	0.8087804913520813
FABULOUS	0.7933964133262634
lovely	0.7787262201309204
fab	0.7482037544250488
fantabulous	0.7478336095809937
amazing	0.7478227615356445

terrific	0.7327693700790405
fabulously	0.7217686772346497
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
.....	0.2282805442810058
.....	0.2280531972646713
.....	0.2278845608234405
.....	0.2249659746885299
■■■■■	0.2249569594860077
.....	0.2247526198625564
.....	0.2247092872858047
.....	0.2241306900978088
Elpaal	0.2231467962265014
.....	0.2227808535099029

Analyzing word: probably	
Top 10 similar words (custom model):	
Similar Word	Similarity
sazerac	0.5383257269859314
proof	0.5245145559310913
irrefutably	0.49512800574302673

project	0.48511847853660583
whoever	0.4807439148426056
probiotic	0.47602611780166626
projection	0.475987046957016
mildly	0.4712304472923279
absurdly	0.4647114872932434
proprietor	0.45897364616394043
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njpjphb	-0.03746243193745613
square	-0.0482802651822567
buche	-0.05016665160655975
shave	-0.05249975249171257
cure	-0.05417898669838905
front	-0.056182343512773514
structure	-0.058063700795173645

pure	-0.06430643051862717
ready	-0.06550032645463943
foot	-0.06640011072158813
Top 10 similar words (pre-trained model):	
Similar Word	Similarity

probaby	0.7859624028205872
probabaly	0.7439422011375427
certainly	0.7379029989242554
Probably	0.7288143038749695
likely	0.7284035682678223
obviously	0.7145461440086365
probalby	0.7134588360786438
might	0.7121331691741943
proably	0.7051565051078796
proabably	0.7050943970680237
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Update0	0.27925950288772583
Events2	0.25915223360061646
SportsMeet	0.25810128450393677

text4	0.25432586669921875
HomeNew	0.2528176009654999
text3	0.2502356171607971
Information3	0.24840788543224335
Services7	0.24761103093624115
2017Night	0.24594424664974213
■■	0.2441820353269577

Analyzing word: accident	
Top 10 similar words (custom model):	
Similar Word	Similarity
accidental	0.8420753479003906
incident	0.8409819006919861
resident	0.736570417881012
dent	0.6663691997528076
presidente	0.646776556968689
rodent	0.6345327496528625
inadequate	0.6195327639579773
identification	0.5875543355941772
opponent	0.5872465968132019
permanent	0.5790309906005859

Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njpjphb	-0.020757704973220825
ginger	-0.06885376572608948
dish	-0.09276743978261948
beverage	-0.1060856357216835
treat	-0.1067635789513588

flavor	-0.10707331448793411
slow	-0.10721052438020706
flavour	-0.112315334379673
vanilla	-0.12104960530996323
nutella	-0.12260131537914276
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
accidents	0.7520097494125366
accident.It	0.7434254288673401
accident.But	0.7423144578933716
accident.The	0.7230600118637085
accident.This	0.7187039852142334
accident.In	0.7080096006393433
acciden	0.7034187912940979

accident.As	0.6994759440422058
accident.A	0.6938973069190979
mishap	0.6933788061141968
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Sapajus	0.20990262925624847
INs	0.20862528681755066

In-Wall	0.20832177996635437
Irides	0.20424972474575043
Sixe	0.20164655148983002
Gay-Friendly	0.19719411432743073
120x120	0.19664080440998077
Teklad	0.19625823199748993
Congregate	0.19332905113697052
GCCs	0.19295267760753632
Analyzing word: ginger	
Top 10 similar words (custom model):	
Similar Word	Similarity
gingersnap	0.844921350479126
ayinger	0.8144975900650024

kenzinger	0.75457364320755
linger	0.7472002506256104
gingerbread	0.7399316430091858
schlesinger	0.7254171967506409
banger	0.6721712350845337
finger	0.6567360162734985
ranger	0.6509155035018921
younger	0.6314328908920288

Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njpjphb	-0.039967577904462814
wawa	-0.05954955145716667
accident	-0.06885375082492828
book	-0.06958667933940887
accidental	-0.07452186197042465
burn	-0.07987704873085022
couple	-0.08024749904870987
issue	-0.08041772991418839
oyster	-0.08186183124780655
find	-0.0887853130698204

Top 10 similar words (pre-trained model):	
Similar Word	Similarity
ginger-	0.7143461108207703
gingery	0.7072740793228149
gingers	0.6843262910842896
ginge	0.6748136281967163
lemongrass	0.6533388495445251
ginger.	0.6480498313903809
cinnamon	0.6350324153900146

coriander	0.6255741119384766
Ginger	0.6237232089042664
lemon-ginger	0.6218132376670837
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Cascio	0.22519038617610931
Landoll	0.20742197334766388
ReviewsOther	0.19843272864818573
Mazzella	0.18960218131542206
RJL	0.18953433632850647
NRVC	0.18471181392669678
Laghi	0.18392731249332428

NetworkView	0.18150204420089722
Battlelines	0.17888842523097992
totalized	0.1767909824848175
Analyzing word: ethnic	
Top 10 similar words (custom model):	
Similar Word	Similarity
geriatric	0.7018923759460449
scenic	0.6638479232788086

iconic	0.6575673818588257
panic	0.649594783782959
icelandic	0.6469981670379639
bariatrics	0.6466595530509949
francisco	0.641343891620636
fishmonger	0.6202296614646912
tectonic	0.6183239817619324
fabric	0.6137653589248657
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
njjpphb	-0.02988250181078911
time	-0.12433119118213654

thank	-0.13493165373802185
take	-0.13561607897281647
year	-0.14933228492736816
amazing	-0.15444648265838623
tavern	-0.1546805500984192
suck	-0.1552506685256958
pizza	-0.15688465535640717
away	-0.15705424547195435
Top 10 similar words (pre-trained model):	

Similar Word	Similarity
ethnical	0.7625183463096619
Ethnic	0.7574706077575684
ethnics	0.7342380285263062
ethinic	0.7249721884727478
non-ethnic	0.7087332606315613
ethnic-religious	0.6979482769966125
ethnicity	0.6923354864120483
ethno-cultural	0.6914461255073547
ethnic-based	0.6810232996940613
ethno-racial	0.6804916262626648

Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
JDLink	0.22449837625026703
FieldView	0.2194172888994217
F-Response	0.21385324001312256
ProMax	0.2127169817686081
182.50	0.21052594482898712
HighSpeed	0.20890174806118011
SSQA	0.2059040665626526
OneShot	0.20511937141418457
ReviewShare	0.20344999432563782

Haselwood	0.20153288543224335
Analyzing word: information	
Top 10 similar words (custom model):	
Similar Word	Similarity
informative	0.8292074203491211
transformation	0.7942317128181458
proclamation	0.7801945805549622
emancipation	0.7518178224563599
rehydration	0.7438330054283142
inform	0.7434272170066833

format	0.7299120426177979
identification	0.7167450189590454
aviation	0.7072336077690125
cancellation	0.7048238515853882
Top 10 dissimilar words (custom model):	
Dissimilar Word	Dissimilarity
bland	-0.01435836497694254
njpjphb	-0.0328696146607399
main	-0.07266328483819962
skip	-0.07677622884511948

dish	-0.081448033452034
black	-0.08765331655740738
usually	-0.08765829354524612
except	-0.0920279324054718
city	-0.09518494457006454
veggie	-0.0958208292722702
Top 10 similar words (pre-trained model):	
Similar Word	Similarity
informaiton	0.7436331510543823
informatin	0.7321422100067139

informatio	0.721504271030426
informati	0.7098613977432251
info	0.7016404867172241
inforamtion	0.7001004219055176
informatioe	0.6844667792320251
informaton	0.6812226176261902
infomation	0.6806102395057678
inofrmation	0.6782036423683167
Top 10 dissimilar words (pre-trained model):	
Dissimilar Word	Dissimilarity
Aronas	0.2167413830757141

Faceplant31	0.21346688270568848
GaryO	0.21050673723220825
AirnationCrew	0.19967682659626007
King4	0.198287695646286
javaj240	0.19668975472450256
mizzdizz	0.19580696523189545
ReplyDeleteCafe	0.19580407440662384
Silverheart	0.19486887753009796
Manliest	0.1941957324743271

Conclusion

- Based on the analysis, the pre-trained model appears to outperform the custom model.
- Possible reasons for this superiority include the larger and more diverse training data used in pre-training (trained on: Common Crawl text and Wikipedia text), which allows the pre-trained model to capture a broader range of linguistic patterns and relationships.
- Additionally, pre-trained models are often fine-tuned on specific tasks, further enhancing their performance for those tasks.