Part-I

We performed two experiments with CBOW. The first experiment consisted of defining our own custom-defined CBOW model, whereas the second one leveraged gensim's Word2Vec. We kept context length as 2 and embedding depth as 100 for all the experiments.

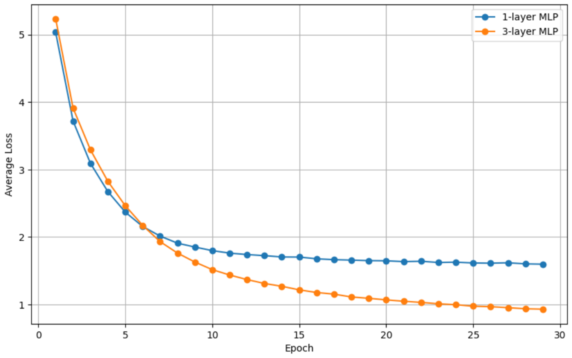


Fig.1. Training Loss for CBOW

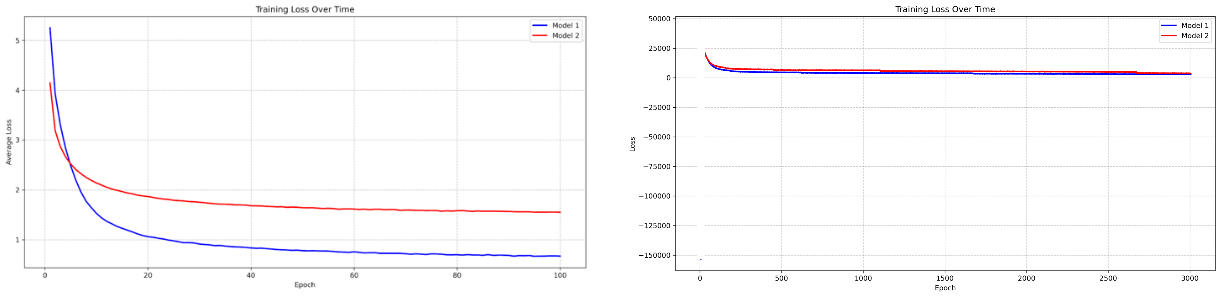
We define disparity (v1, v2) = abs (cos ((v1, E1), (v2, E1)) - cos ((v1, E2), (v2, E2))), and consider only for those cases where cos ((v1, E1), (v2, E1)) * cos ((v1, E2), (v2, E2)) < 0.

Disparity statistics for the two CBOWs:

CBOW	Min	Max	Mean	Std. Deviation
Custom	0.5290	0.7460	0.5745	0.0416
gensim	0.4641	0.9442	0.9442	0.0699

Number of common token pairs in top-100 token disparity pairs: 0

CBOW	Pair with highest disparity
Custom	'old' and 'expect'
gensim	'tal' and 'bank'



(a)

(b)

Fig. 2. Training loss with (a) custom CBOW (b) gensim CBOW for legal document (Model 1) and literature document (Model 2)

Disparity statistics between CBOW and Skip-Gram:

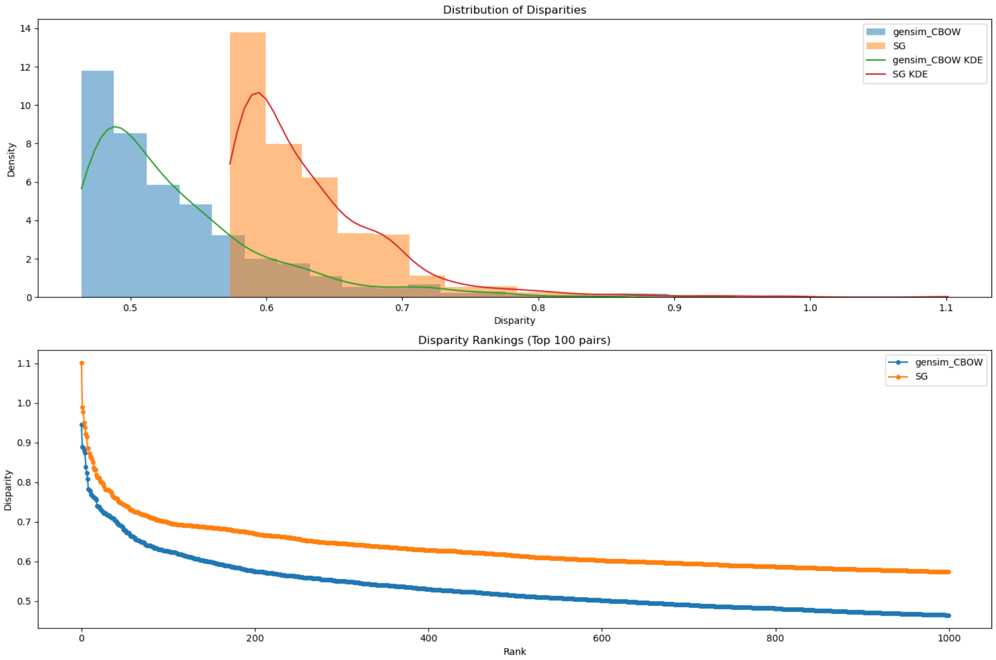


Fig. 3. Disparity comparison between gensim CBOW and Skip-Gram for top-1000 pairs

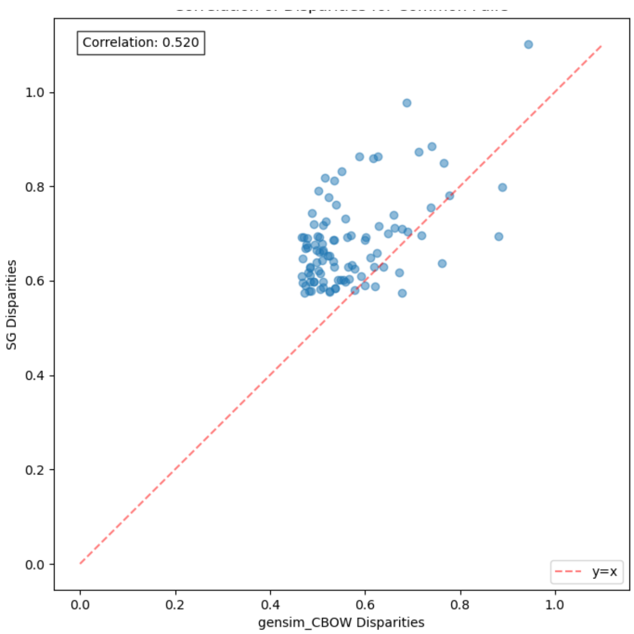


Fig. 4. Disparity correlation for common pairs in top disparities

Part-II

Number of citations in the document “Model-output.txt” = 11
Number of Type 1 Hallucinations = 10

S. No.	LLM-cited reference	Original reference (closest match)	Ground Truth	Cosine Similarity
1	Hurwitz v. United States, 884 F.2d 684, 687 (2d Cir. 1989)	Hurwitz v. United States, 208 F. Supp. 594 (S.D. Tex. 1962)	False	0.2944
2	Messenger v. Gruner Key Symbol Jahr Printing and Publ’g, 94 N.Y.2d 436, 441, 706 N.Y.S.2d 52, 54 (2000)	Messenger v. Gruner + Jahr Printing & Publishing, 727 N.E.2d 549 (NY 2000) OR Messenger v Gruner + Jahr Print. & Publ., 94 NY2d 436, 441 [2000]	True	0.38
3	D'Andrea v. Fakename, 972 F.Supp. 154, 157 (E.D.N.Y. 1997)	D'ANDREA v. Rafla-Demetrious, 972 F. Supp. 154 (E.D.N.Y 1997)	True	0.442
4	Weil v. Johnson, Index No. 119431/02, 2002 WL 31972157, *4-5 (Sup.Ct. N.Y. Co. Sept 27, 2002)	Johnston v. Weil, 946 N.E.2d 329 (Ill. 2011)	False	0.2304
5	Piskac v. Shapiro, 230 Conn. 345 (2025)	Shapiro v. Thompson, 394 U. S. 618, 634 (1969)	N/A	N/A
6	Lemerond v. Twentieth Century Fox Film Corp., 564 F.Supp.2d 315, 323 (S.D.N.Y.)	Polydoros v. Twentieth Century Fox Film Corp., 67 Cal. App. 4th 318, 79 Cal. Rptr. 2d 207 (1997)	Somewhat True	0.3479
7	Delan by Delan v. CBS, Inc., 91 A.D.3d 255, 458 N.Y.S.23d 608 (2d Dep’t 2013)	Delan by Delan v. CBS, Inc., 91 A.D.2d 255, 458 N.Y.S.2d 608, 614 (2d Dep't 1983)	Somewhat True	0.3760
8	Finger v. Omni Publs. Intl.,	Finger v. Omni Publications	True	0.5049

	Ltd., 157 A.D.2d 956, 964, 77 N.Y.S.24d 138, 141 (4th Dep’t 1990)	International., Ltd., 77 N.Y.2d 138 (NY 1990)		
9	Arrington v. New York Times Co., 55 N.Y.22d 433, 440, 449 N.Y.S.22d 941, 944, 434 N.E.22d 1319, 1322 (1982)	Arrington v. New York Times Co., 55 N.Y.2d 433, 439, 449 N.Y.S.2d 941, 434 N.E.2d 1319 (1982)	True	0.3894
10	Spurlock v. Candelaria, 08 Civ. 1830 (BMC) (RER), E.D.N.Y. Jul. 3, 2008)	N/A	N/A	N/A
11	Gautier v. Pro-Football, Inc., 304 N.Y. 354, 107 N.E.2d 485 (1952)	Gautier v. Pro-Football, Inc., 304 N.Y. 354, 359, 107 N.E.2d 485 (1952) AND Gautier v. Pro-Football, 304 N.Y. 354, 359, supra AND Gautier v. Pro-Football, Inc., 107 N.E.2d 485 (NY 1952)	True	0.7473

Observations:

1. It hallucinates often with volume number ([1]) and adds an incorrect parallel citation ([2], 706 N.Y.S.2d 52, 54; [7], 614)
2. It places an incorrect plaintiff’s or defendant’s name ([3, 4, 5, 6])
3. It mismatches the series number the most ([7], 2d instead of 3d, 2d instead of 23d; [8] 2d instead of 24d; [9] 2d instead of 22d).
4. Rarely mismatches dates ([7], [4]). An interesting thing about [4] is the presence of the followibg lines in the database: “Johnston subsequently married Andrew Weil and, in 2002, they had a daughter.” This probably influenced it to use 2002 instead of original year 2011 (ignoring the typo of Johnston to Johnson).
5. Since we did exact matching between LLM-cited references and the original ones, citation [11] is controversial to claim as Type 1 hallucination. It missed the volume no. 359 but it still referenced the correct citation in my opinion.

Hence, we do a cosine similarity analysis between the contexts for citation [11] in the database and the LLM response. The contexts are as follows:

sentence1 = "Any person whose name, portrait or picture is used within this state for advertising purposes or for the purposes of trade without [such person's] written consent * may * * *sue and recover damages for any injuries sustained by reason of such use" #from database sentence2 = "Accordingly, if a person brings a claim against a documentary filmmaker for the use of their image in the documentary, the claim would likely fail because a documentary is not deemed produced for the purposes of advertising or trade" #from LLM response

Cosine similarity of their embedding vectors = 0.4758

We also show it for the other citations in the same table for a comprehensive comparison. We conclude that the context size is very important, and a particular section of the context (where rules or laws are concluded by the courts) are more important than the intricate details (about persons, characters. Place, etc.). More or less, based on my readings, I label the citations as true or false based on whether the LLM response uses them correctly in the arguments. Cosine similarity trends show score > 0.38 to be correct in terms of the context, while lower scores (< 0.3) are in accordance with incorrect context.