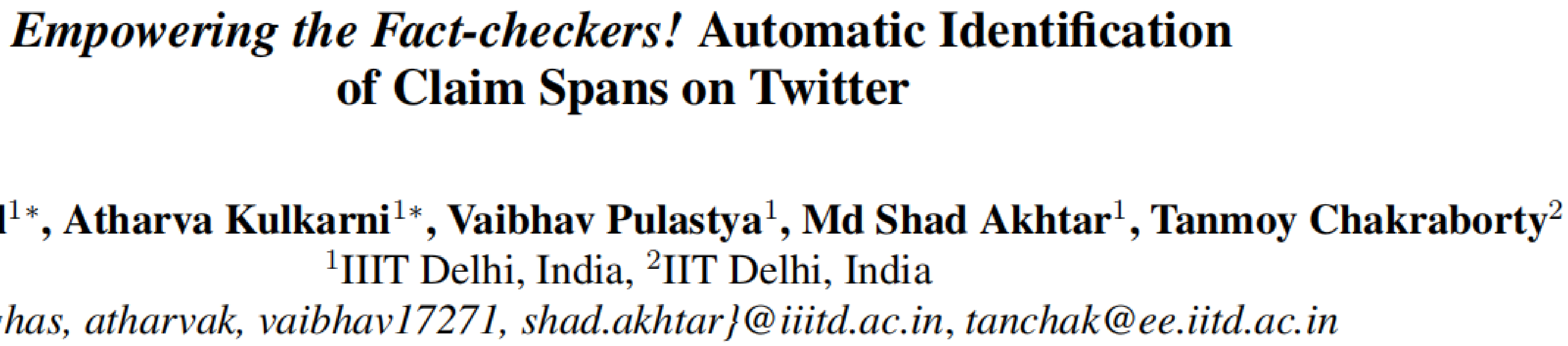


Empowering the Fact-checkers! Automatic Identification of Claim Spans on Twitter

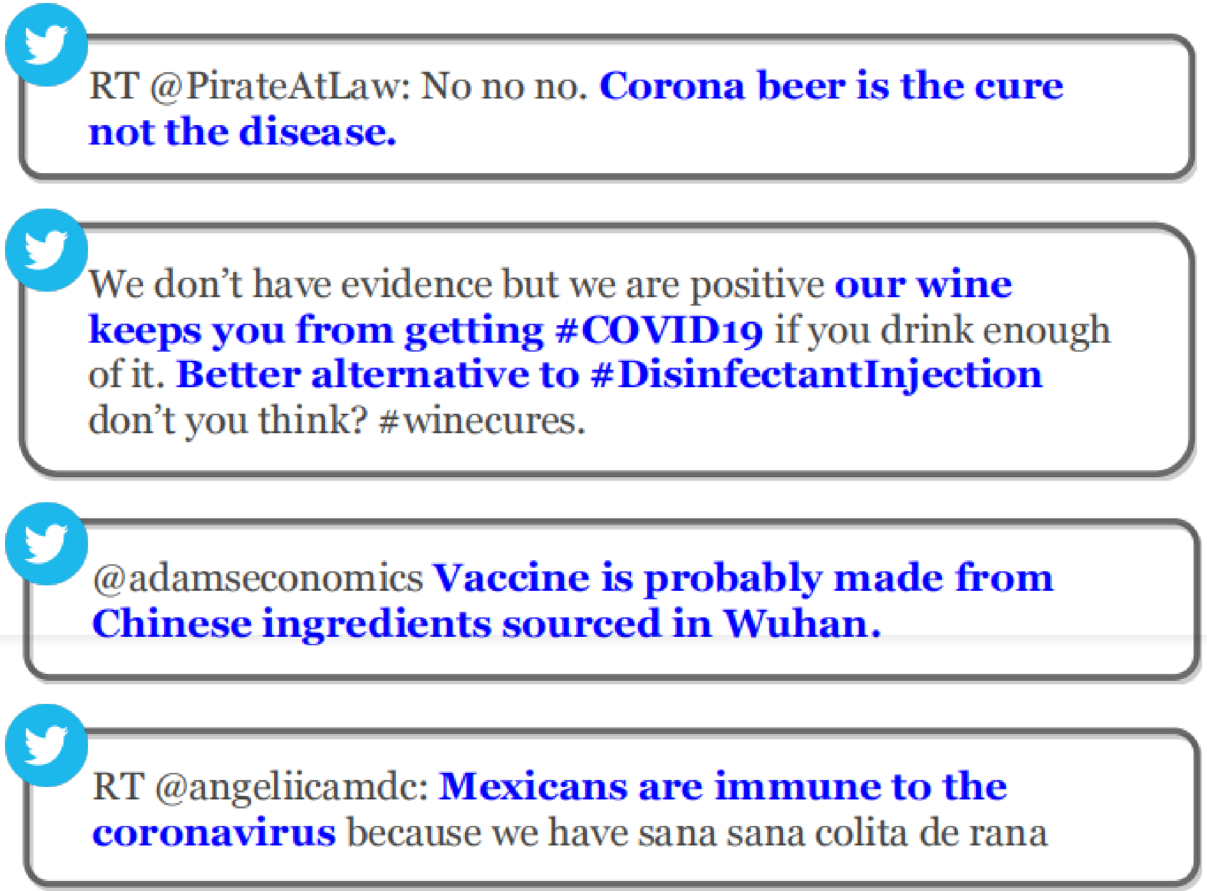
Megha Sundriyal[1*], Atharva Kulkarni[1*], Vaibhav Pulastya[1], Md Shad Akhtar[1], Tanmoy Chakraborty[2]
[1]IIIT Delhi, India, [2]IIT Delhi, India
{meghas, atharvak, vaibhav17271, shad.akhtar}@iiitd.ac.in, tanchak@ee.iitd.ac.in

code:https://github.com/LCS2-IIITD/DABERTA-EMNLP-2022.

EMNLP_2022

Reported by Xiaoke Li

Introduction



RT @PirateAtLaw: No no no. **Corona beer is the cure not the disease.**

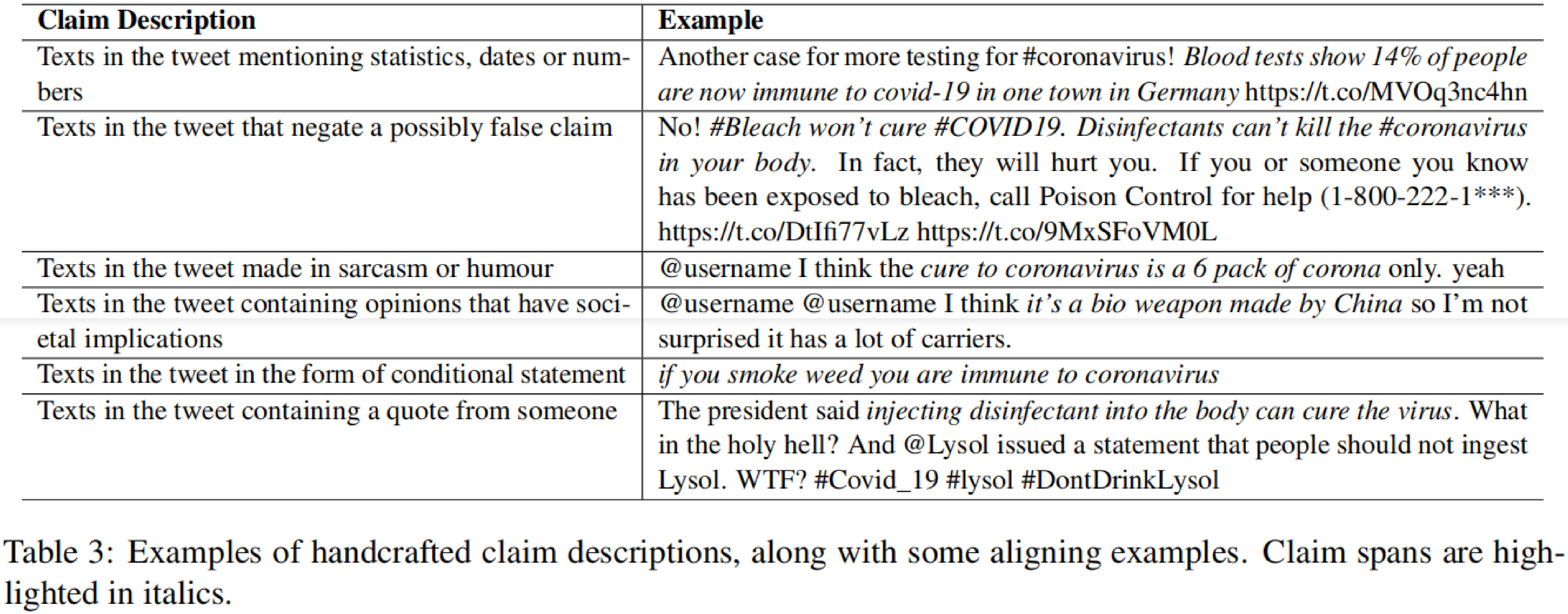
We don’t have evidence but we are positive **our wine keeps you from getting #COVID19** if you drink enough of it. **Better alternative to #DisinfectantInjection** don’t you think? #winecures.

@adamseconomics **Vaccine is probably made from Chinese ingredients sourced in Wuhan.**

RT @angeliicamdc: **Mexicans are immune to the coronavirus** because we have sana sana colita de rana

Figure 1: Examples of claim tweets and their ground truth claim spans highlighted in boldface text (blue).

Method



Claim Description	Example
Texts in the tweet mentioning statistics, dates or numbers	Another case for more testing for #coronavirus! *Blood tests show 14% of people are now immune to covid-19 in one town in Germany* https://t.co/MVOq3nc4hn
Texts in the tweet that negate a possibly false claim	No! *#Bleach won't cure #COVID19. Disinfectants can't kill the #coronavirus in your body*. In fact, they will hurt you. If you or someone you know has been exposed to bleach, call Poison Control for help (1-800-222-1***). https://t.co/DtIfi77vLz https://t.co/9MxSFoVM0L
Texts in the tweet made in sarcasm or humour	@username I think the *cure to coronavirus is a 6 pack of corona* only. yeah
Texts in the tweet containing opinions that have societal implications	@username @username I think *it's a bio weapon made by China* so I'm not surprised it has a lot of carriers.
Texts in the tweet in the form of conditional statement	*if you smoke weed you are immune to coronavirus*
Texts in the tweet containing a quote from someone	The president said *injecting disinfectant into the body can cure the virus*. What in the holy hell? And @Lysol issued a statement that people should not ingest Lysol. WTF? #Covid_19 #lysol #DontDrinkLysol

Table 3: Examples of handcrafted claim descriptions, along with some aligning examples. Claim spans are highlighted in italics.

Method

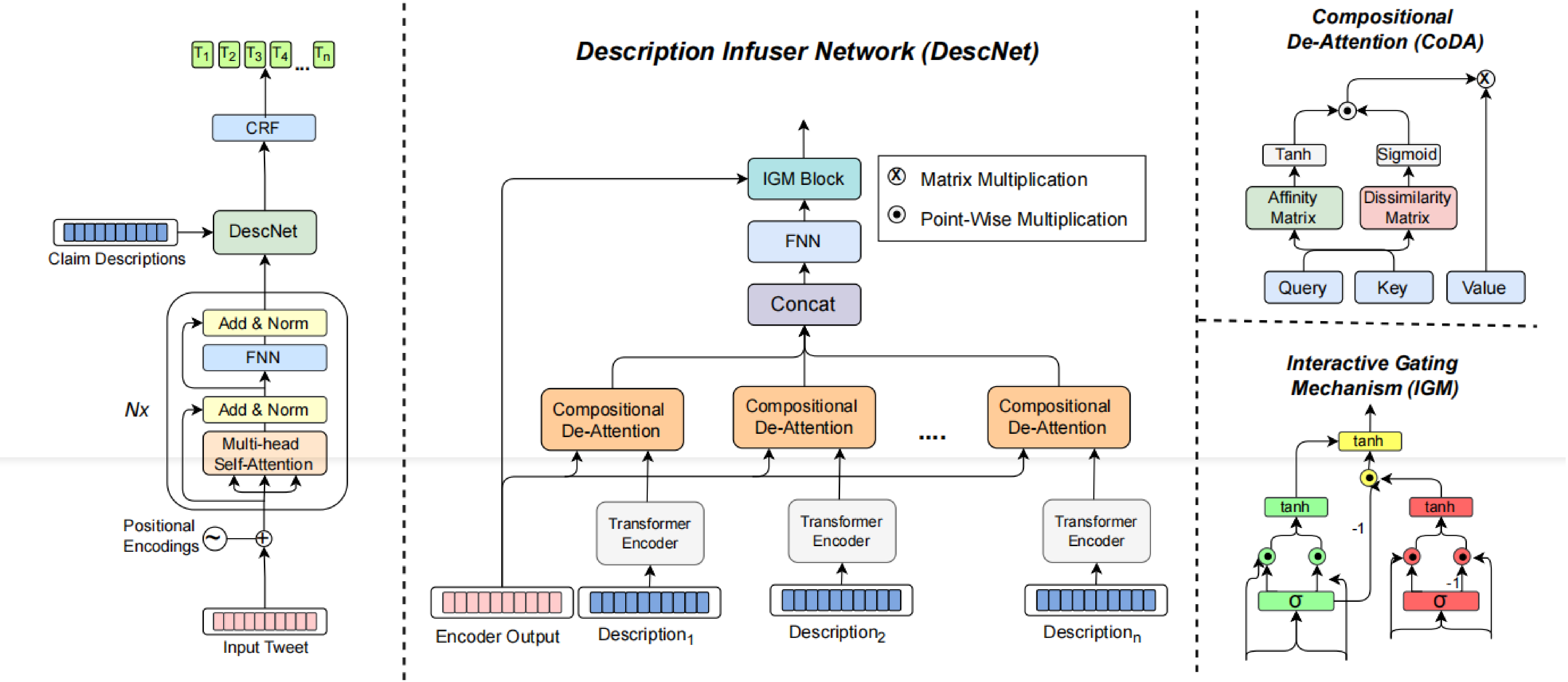
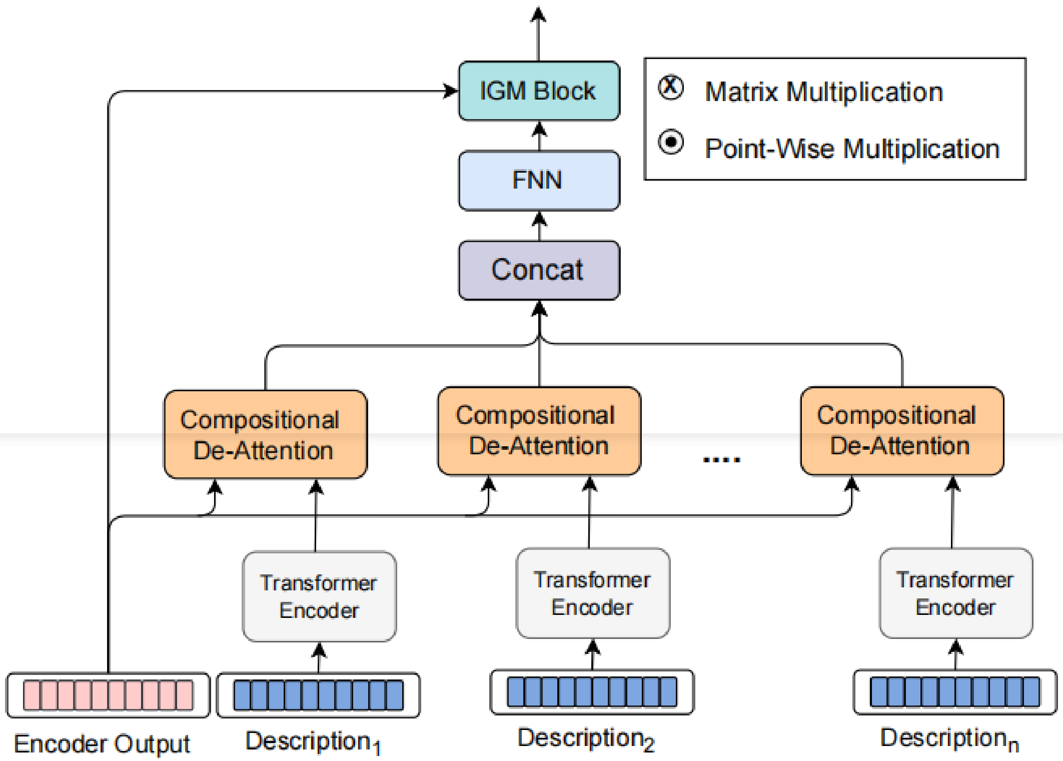


Figure 2: A schematic diagram of DABERTa for the claim span identification. ⊙ represents point-wise multiplication, and ⊗ represents matrix multiplication.

Method

Description Infuser Network (DescNet)

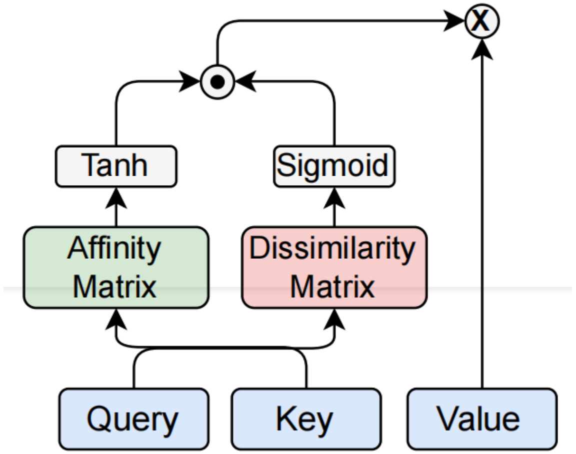


$$Z_{ij}^{C} = CoDA(Z_i, D_j)D_j \qquad (1)$$

$$Z_i' = Concat(Z_{i1}^{C}, ..., Z_{im}^{C}) \qquad (2)$$

$$\hat{Z}_i = IGM(Z_i'W, Z_i) \qquad (3)$$

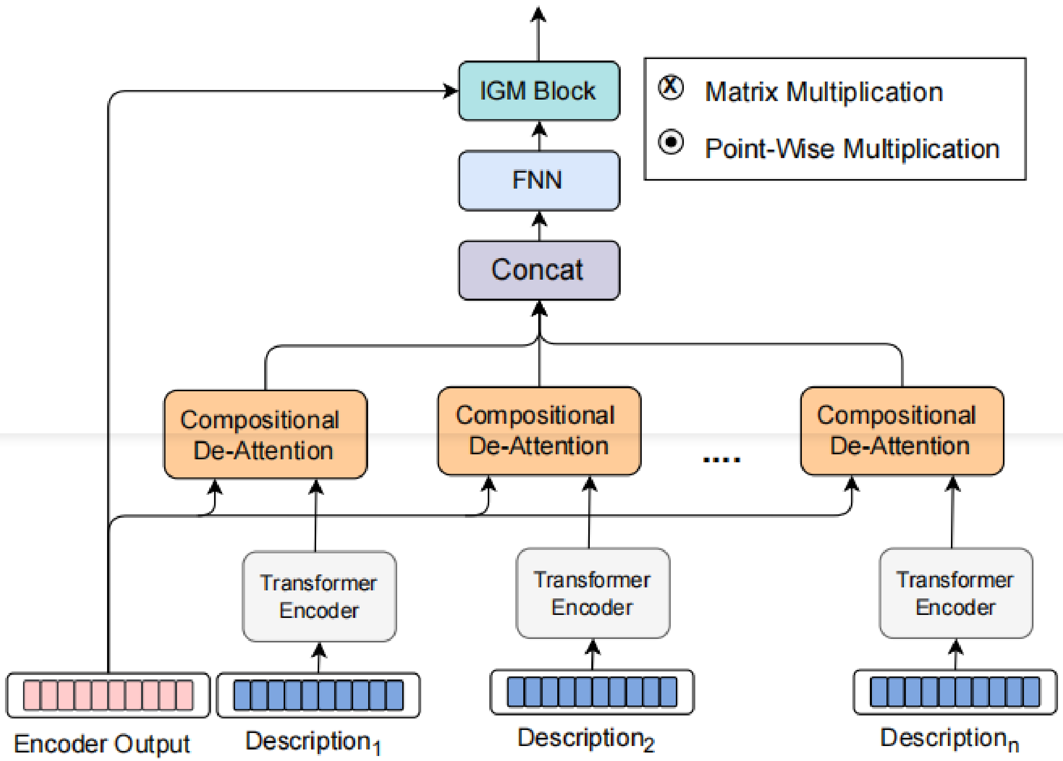
Compositional De-Attention (CoDA)



$$A_{quasi} = \left(tanh(\frac{QK^T}{\sqrt{d_k}}) \odot \sigma(\frac{G(Q,K)}{\sqrt{d_k}}) \right) V \qquad (4)$$

Method

Description Infuser Network (DescNet)

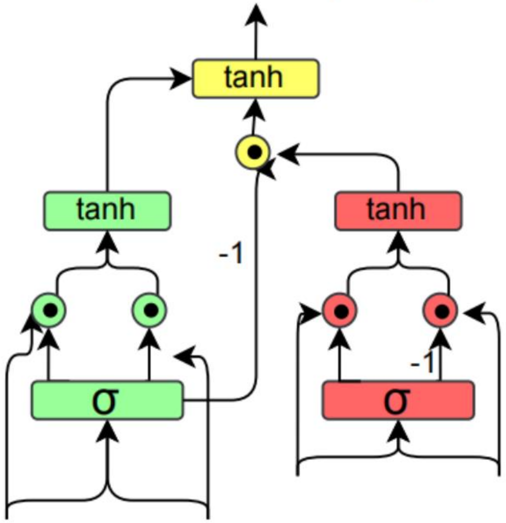


$$A = R + (1 - \mu_r) \odot C \tag{9}$$

$$\hat{Z}_i = tanh(AW_a + b_a) \odot Z_i \tag{10}$$

Finally, this vector $\hat{Z}_i$ is passed to a CRF layer for token classification.

Interactive Gating Mechanism (IGM)



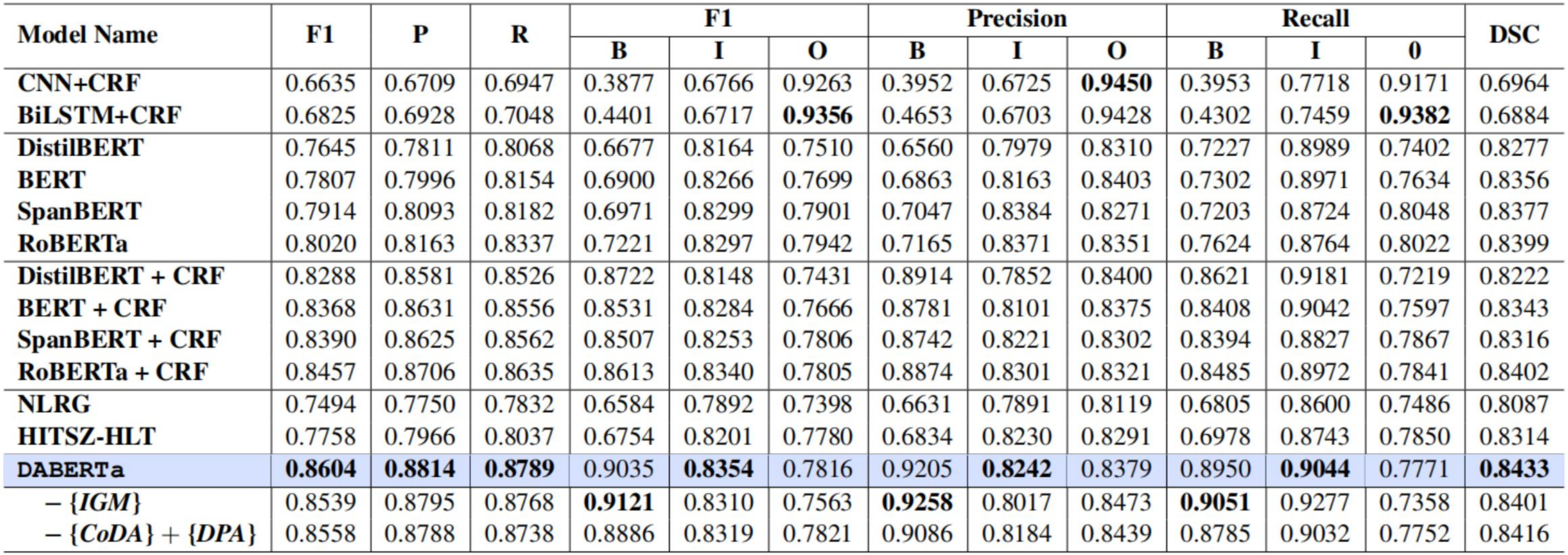
$$\mu_c = \sigma(Z_{ip}W_{c1} + Z'_{ip}W_{c2} + b_{c1}) \tag{5}$$

$$C = \tanh(Z_{ip} \odot \mu_c W_{c3} + Z'_{ip} \odot (1 - \mu_c)W_{c4} + b_{c2}) \tag{6}$$

$$\mu_r = \sigma(Z_{ip}W_{r1} + Z'_{ip}W_{r2} + b_{r1}) \tag{7}$$

$$R = \tanh(Z_{ip} \odot \mu_r W_{r3} + Z'_{ip} \odot \mu_r W_{r4} + b_{r2}) \tag{8}$$

Experiments



Model Name	F1	P	R	F1			Precision			Recall			DSC
				B	I	O	B	I	O	B	I	0	
CNN+CRF	0.6635	0.6709	0.6947	0.3877	0.6766	0.9263	0.3952	0.6725	**0.9450**	0.3953	0.7718	0.9171	0.6964
BiLSTM+CRF	0.6825	0.6928	0.7048	0.4401	0.6717	**0.9356**	0.4653	0.6703	0.9428	0.4302	0.7459	**0.9382**	0.6884
DistilBERT	0.7645	0.7811	0.8068	0.6677	0.8164	0.7510	0.6560	0.7979	0.8310	0.7227	0.8989	0.7402	0.8277
BERT	0.7807	0.7996	0.8154	0.6900	0.8266	0.7699	0.6863	0.8163	0.8403	0.7302	0.8971	0.7634	0.8356
SpanBERT	0.7914	0.8093	0.8182	0.6971	0.8299	0.7901	0.7047	0.8384	0.8271	0.7203	0.8724	0.8048	0.8377
RoBERTa	0.8020	0.8163	0.8337	0.7221	0.8297	0.7942	0.7165	0.8371	0.8351	0.7624	0.8764	0.8022	0.8399
DistilBERT + CRF	0.8288	0.8581	0.8526	0.8722	0.8148	0.7431	0.8914	0.7852	0.8400	0.8621	0.9181	0.7219	0.8222
BERT + CRF	0.8368	0.8631	0.8556	0.8531	0.8284	0.7666	0.8781	0.8101	0.8375	0.8408	0.9042	0.7597	0.8343
SpanBERT + CRF	0.8390	0.8625	0.8562	0.8507	0.8253	0.7806	0.8742	0.8221	0.8302	0.8394	0.8827	0.7867	0.8316
RoBERTa + CRF	0.8457	0.8706	0.8635	0.8613	0.8340	0.7805	0.8874	0.8301	0.8321	0.8485	0.8972	0.7841	0.8402
NLRG	0.7494	0.7750	0.7832	0.6584	0.7892	0.7398	0.6631	0.7891	0.8119	0.6805	0.8600	0.7486	0.8087
HITSZ-HLT	0.7758	0.7966	0.8037	0.6754	0.8201	0.7780	0.6834	0.8230	0.8291	0.6978	0.8743	0.7850	0.8314
DABERTa	**0.8604**	**0.8814**	**0.8789**	0.9035	**0.8354**	0.7816	0.9205	**0.8242**	0.8379	0.8950	**0.9044**	0.7771	**0.8433**
– {*IGM*}	0.8539	0.8795	0.8768	**0.9121**	0.8310	0.7563	**0.9258**	0.8017	0.8473	**0.9051**	0.9277	0.7358	0.8401
– {*CoDA*} + {*DPA*}	0.8558	0.8788	0.8738	0.8886	0.8319	0.7821	0.9086	0.8184	0.8439	0.8785	0.9032	0.7752	0.8416

Table 4: Experimental results of DABERTa, its variants (last two rows), and baselines. DSC, P, and R denote Dice Similarity Coefficient, Precision, and Recall respectively.

Experiments

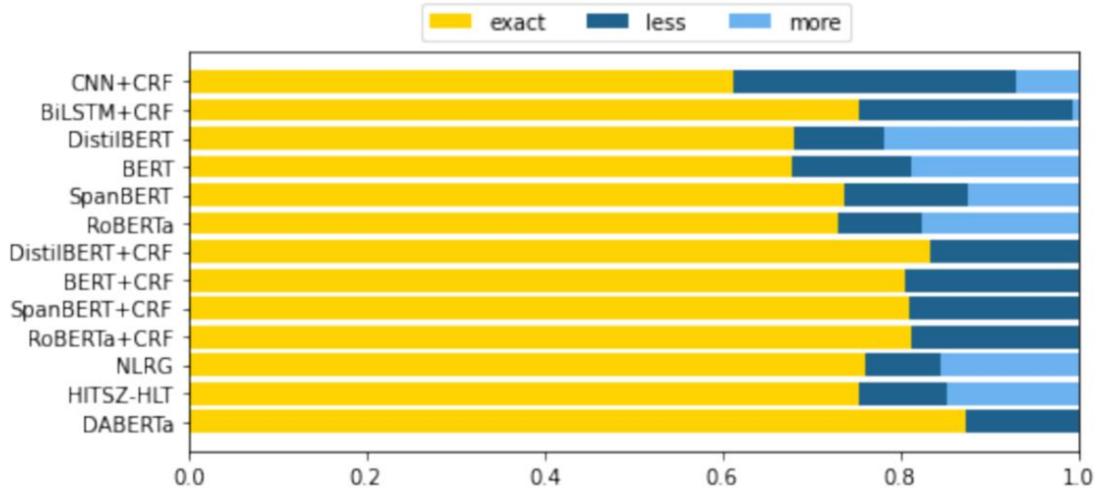


Figure 3: A comparative study among DABERTa and baselines. The horizontal bar signifies the ration of number of predicted spans and number of gold spans.

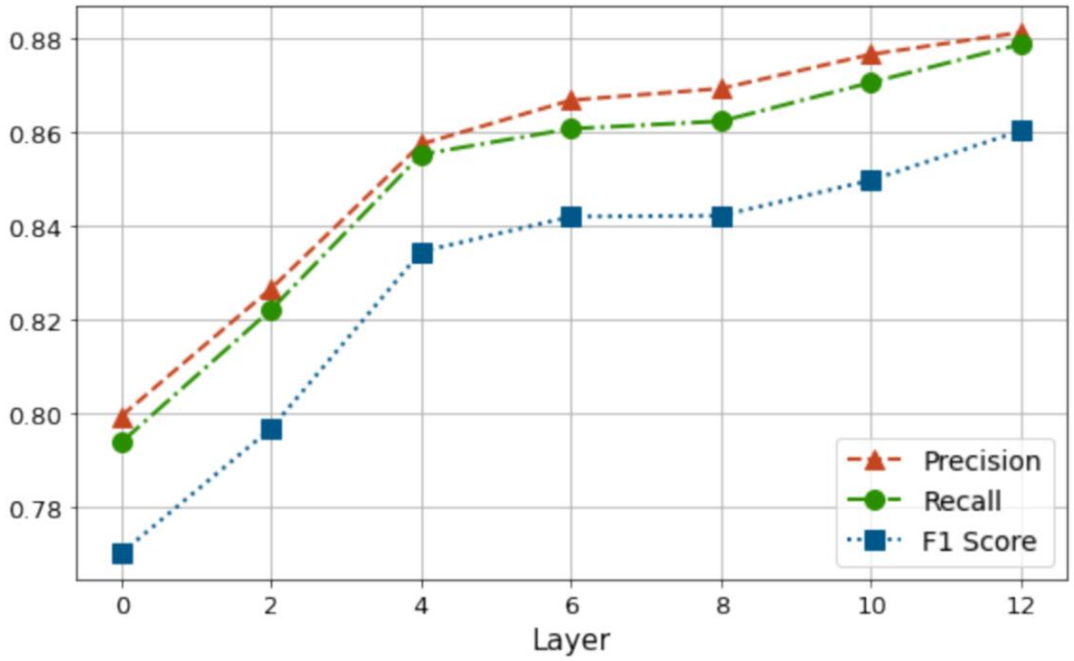
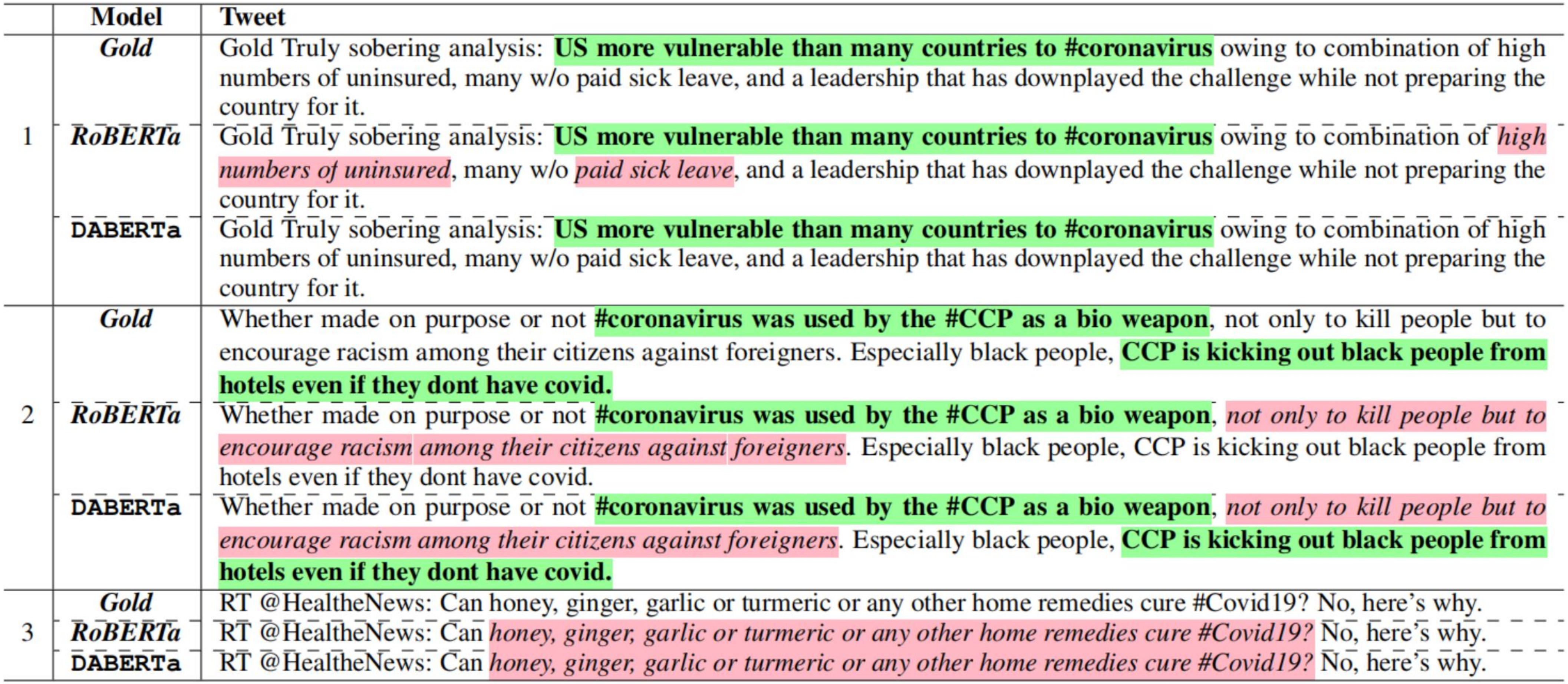


Figure 4: Performance of DABERTa when the adapter module is inserted at different layers of RoBERTa.

Experiments



	Model	Tweet
1	***Gold***	Gold Truly sobering analysis: **US more vulnerable than many countries to #coronavirus** owing to combination of high numbers of uninsured, many w/o paid sick leave, and a leadership that has downplayed the challenge while not preparing the country for it.
	RoBERTa	Gold Truly sobering analysis: **US more vulnerable than many countries to #coronavirus** owing to combination of *high numbers of uninsured*, many w/o *paid sick leave*, and a leadership that has downplayed the challenge while not preparing the country for it.
	DABERTa	Gold Truly sobering analysis: **US more vulnerable than many countries to #coronavirus** owing to combination of high numbers of uninsured, many w/o paid sick leave, and a leadership that has downplayed the challenge while not preparing the country for it.
2	***Gold***	Whether made on purpose or not **#coronavirus was used by the #CCP as a bio weapon**, not only to kill people but to encourage racism among their citizens against foreigners. Especially black people, **CCP is kicking out black people from hotels even if they dont have covid.**
	RoBERTa	Whether made on purpose or not **#coronavirus was used by the #CCP as a bio weapon**, *not only to kill people but to encourage racism among their citizens against foreigners*. Especially black people, CCP is kicking out black people from hotels even if they dont have covid.
	DABERTa	Whether made on purpose or not **#coronavirus was used by the #CCP as a bio weapon**, *not only to kill people but to encourage racism among their citizens against foreigners*. Especially black people, **CCP is kicking out black people from hotels even if they dont have covid.**
3	***Gold***	RT @HealtheNews: Can honey, ginger, garlic or turmeric or any other home remedies cure #Covid19? No, here’s why.
	RoBERTa	RT @HealtheNews: Can *honey, ginger, garlic or turmeric or any other home remedies cure #Covid19?* No, here’s why.
	DABERTa	RT @HealtheNews: Can *honey, ginger, garlic or turmeric or any other home remedies cure #Covid19?* No, here’s why.

Table 5: Error analysis of the outputs. Bold text (green) highlights the correct claim span whereas text in italics (red) represents the mistakes committed by our model, DABERTa, and vanilla RoBERTa as baseline.

Thanks