

TOPIC: Claim2Source at CheckThat! 2025:
Zero-Shot Style Transfer for Scientific Claim-Source Retrieval

SPEAKER: Tobias Schreieder



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.

CheckThat! 2025 Task 4b

Scientific Claim Source Retrieval

Claim

Myth: I sometimes see the downplaying of COVID deaths because those dying "are old" on average and would have "died soon anyway". **This study looked at the first 1,279,866 reported #covid19 deaths.** The estimated average years of life lost was 16 years

retrieve

Source

Years of life lost to COVID-19 in 81 countries. Understanding the mortality impact of COVID-19 requires not only counting the dead, but analyzing how premature the deaths are. We calculate **years of life lost (YLL)** across 81 countries due to COVID-19 attributable deaths, and also conduct an analysis based on estimated excess deaths. We find that over 20.5 million years of life have been lost to COVID-19 globally. As of January 6, 2021, YLL in heavily affected countries are 2–9 times the average seasonal influenza; three quarters of the YLL result from deaths in ages below 75 and almost a third from deaths below 55; and men have lost 45% more life years than women. The results confirm the large mortality impact of COVID-19 among the elderly. They also call for heightened awareness in devising policies that protect vulnerable demographics losing the largest number of life-years.

Goal: Identify the scientific paper referenced in a social media post

- **Claims:** Tweets containing scientific findings expressed in everyday language
- **Sources:** Titles, abstracts, and metadata from COVID-19 research papers (CORD-19)

Task: Match each tweet to the original scientific paper it refers to

Motivation – Style of Texts

Claim

Myth: I sometimes see the downplaying of COVID deaths because those dying "are old" on average and would have "died soon anyway". **This study looked at the first 1,279,866 reported #covid19 deaths.** The estimated average years of life lost was 16 years

Source

Years of life lost to COVID-19 in 81 countries. Understanding the mortality impact of COVID-19 requires not only counting the dead, but analyzing how premature the deaths are. We calculate **years of life lost (YLL)** across 81 countries due to COVID-19 attributable deaths, and also conduct an analysis based on estimated excess deaths. We find that over 20.5 million years of life have been lost to COVID-19 globally. As of January 6, 2021, YLL in heavily affected countries are 2–9 times the average seasonal influenza; three quarters of the YLL result from deaths in ages below 75 and almost a third from deaths below 55; and men have lost 45% more life years than women. The results confirm the large mortality impact of COVID-19 among the elderly. They also call for heightened awareness in devising policies that protect vulnerable demographics losing the largest number of life-years.

Style of Claim

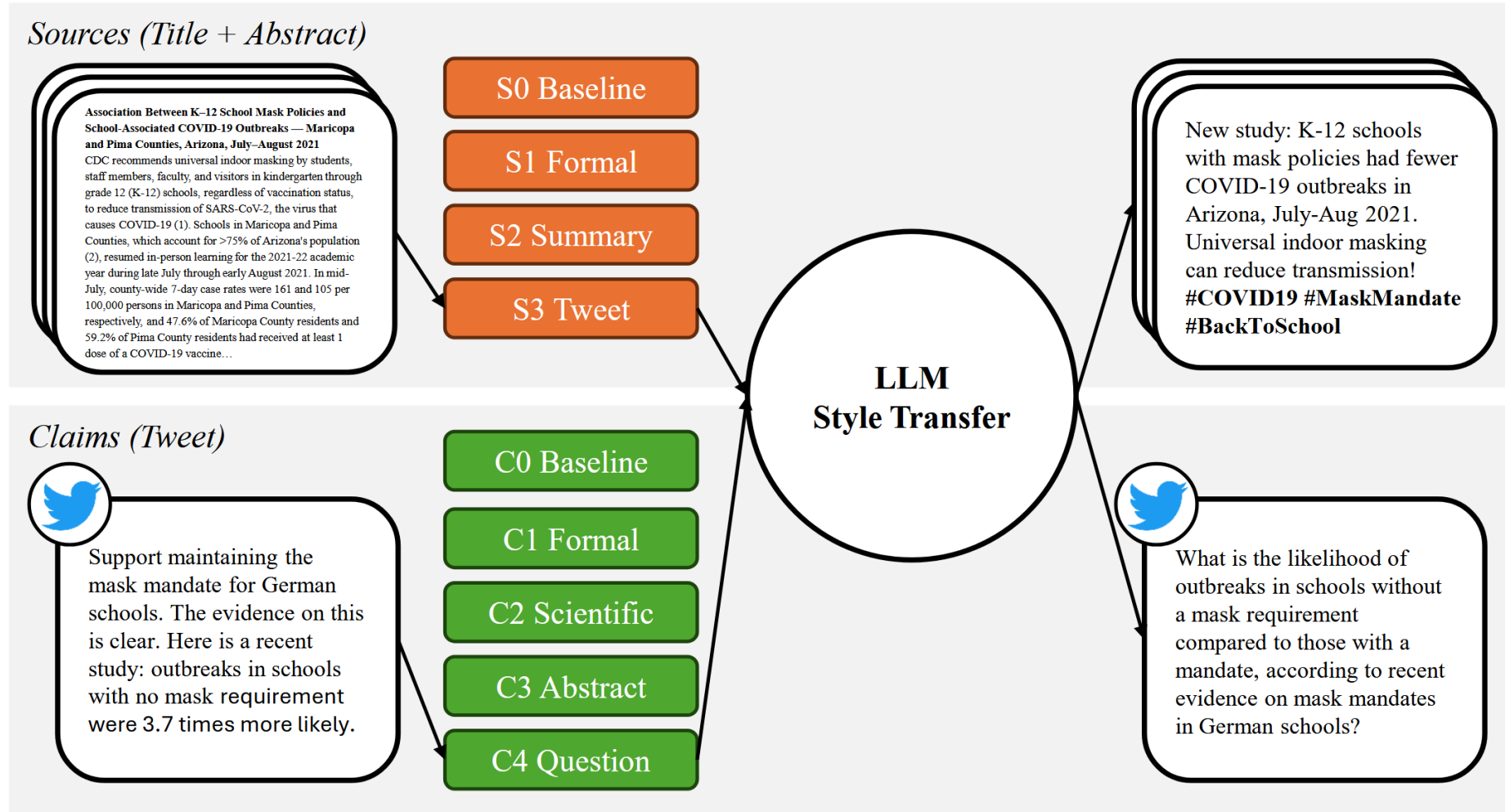
- Short & concise
- Informal tone
- Everyday, general language



Style of Source

- Long & detailed
- Formal, academic tone
- Technical, scientific language

Text Style Transfer - Pipeline



Text Style Transfer - Implementation

- Zero-shot prompts using **LLaMA 3.3 70B Instruct**
- **9 different text styles:**
 - **5 styles for claims** (*incl. original style*)
 - **4 styles for sources** (*incl. original style*)
- **Modular Prompting Template:**
 - **Context:** Defines overall goal of claim-source retrieval with style transfer
 - **Task:** Specifies LLM's task
 - **Instructions:** Guides styles transfer
 - **Output specifications:** Ensures consistent format

Modular Prompting Template

➡ **Context:** You are a language model specialized in text optimization for scientific claim-source retrieval. In this task informal social media posts like tweets (claim) have to be matched with the most relevant scientific abstract (source). To increase the retrieval performance, a style transfer of the given tweet must be performed.

➡ **Task:** Pre-process the given tweet with minimal changes, according to the following instructions.

➡ **Instructions:**

- Remove non-essential elements like hashtags, emojis, and informal symbols.
- Retain all key details, and all factual information.
- Do not alter the terminology or the core meaning of the tweet.
- Focus on improving readability without over-processing, ensuring the tweet remains suitable for matching with corresponding scientific papers.

C1

Task: Rewrite each tweet into a more formal scientific-claim format, according to the following instructions.

Instructions:

- Transform the tweet into a clear, standalone scientific claim or set of claims.
- Remove non-essential elements like hashtags, emojis, and informal symbols.
- Do not alter the meaning of the tweet.
- Use precise academic language and domain-specific terminology. Do not alter the terminology of the tweet, if it is already scientific.

C2

Task: For each tweet write a scientific abstract, according to the following instructions.

Instructions:

- Write a concise, formal scientific abstract (~150 words).
- Use precise academic language and domain-specific terminology. Do not alter the terminology of the tweet, if this is already scientific.
- Ensure coherence and logical flow, but do not include section headers.
- Present key aspects of the study, including context, methodology (if available), key findings, and significance, in a continuous and structured manner. Do not introduce information which is not present in the tweet.
- Maintain neutrality unless the tweet explicitly expresses a stance, in which case incorporate the sentiment appropriately.

C3

Task: For each tweet write one scientific question, according to the following instructions.

Instructions:

- The question should only be answerable with the paper the tweet refers to.
- The question should be formulated precisely without being divided into several sub-questions.
- Do not alter the terminology used in the tweet.
- The question should be optimized for the best possible retrieval performance.

C4

➡ **Output:** Only return the pre-processed question without any explanations or step-by-step reasoning!

Text Styles - Claims

Myth: I sometimes see the downplaying of COVID deaths because those dying are old on average and would have died soon anyway. This study looked at the first 1,279,866 reported covid19 deaths. The estimated average years of life lost was 16 years.

C1 Formal

- Rewrite tweet into formal tone
- Improve readability: remove hashtags, emojis
- Preserve terminology and meaning

A common misconception surrounding COVID-19 is that the deaths are often downplayed due to the advanced age of the individuals, with the implication that they would have died soon regardless. However, a study examining the first 1,279,866 reported COVID-19 deaths found that the estimated average years of life lost was 16 years.

C2 Scientific

- Rewrite tweet into standalone scientific claims
- Use precise academic language & domain-specific terminology
- Preserve meaning and don't alter existing scientific terminology

Text Styles - Claims

This study examined the first 1,279,866 reported COVID-19 deaths, aiming to contextualize the impact of the pandemic on mortality. A common misconception downplays the significance of COVID-19-related deaths, attributing them to the advanced age of the deceased and implying that these individuals would have succumbed to other causes shortly thereafter. However, the findings of this investigation reveal a substantial loss of life expectancy, with an estimated average of 16 years of life lost per deceased individual. This metric underscores the considerable premature mortality associated with COVID-19, highlighting the need to reevaluate the perceived insignificance of these deaths. The results emphasize the profound consequences of the pandemic, particularly in terms of the years of life forfeited, and underscore the importance of acknowledging the true extent of COVID-19's impact on human lifespan.

C3 Abstract

- Write a concise scientific abstract of ~150 words
- Present key aspects of the study, including context, methodology, key findings, ... in a structured manner
- Maintain neutrality unless the tweet explicitly expresses a stance

What is the estimated average years of life lost among the first 1,279,866 reported COVID-19 deaths?

C4 Question

- Rewrite tweet into a single scientific question, without being divided into sub-questions
- Preserve terminology and meaning
- Optimize question for the best possible retrieval performance

Text Styles - Sources

To understand the mortality impact of COVID-19, it's essential to analyze the prematurity of deaths, not just the number of deaths. We calculated years of life lost (YLL) due to COVID-19 in 81 countries and found that over 20.5 million years of life have been lost globally. As of January 6, 2021, heavily affected countries experienced 2-9 times more YLL than the average seasonal influenza. Notably, three-quarters of YLL resulted from deaths in people under 75, and almost a third from deaths in people under 55. Men lost 45% more life years than women. The results highlight the significant mortality impact of COVID-19, particularly among the elderly and other vulnerable demographics, emphasizing the need for policies that protect these groups.

S1 Formal

- Remove unnecessary filler or verbose academic phrasing, while preserving terminology and meaning
- Fix unclear grammar or overly long sentences for readability
- Omit references to the paper itself, if not essential to meaning

We calculated years of life lost (YLL) due to COVID-19 in 81 countries, finding over 20.5 million years lost globally. Analysis revealed that YLL in heavily affected countries are 2-9 times higher than average seasonal influenza, with 75% of YLL resulting from deaths under 75 years and 30% from deaths under 55. Men lost 45% more life years than women, highlighting the significant mortality impact of COVID-19, particularly among the elderly and younger demographics.

S2 Summary

- Summarize the core findings, relevant methodology and key contextual details, while preserving terminology and meaning
- Focus on content that supports scientific claims and aids in source retrieval.
- Exclude general background, broad motivation, or unrelated information

Text Styles - Sources

"COVID-19 has claimed 20.5M+ years of life globally! 75% of years lost are from people under 75. Men lost 45% more years than women. #COVID19 #GlobalHealth #MortalityRate"

S3 Tweet

- Write a concise tweet (max. 280 characters) that draws attention to the key research findings
- The tweet should be suitable for both the paper's author and third parties to share
- Use relevant hashtags to increase discoverability

Retrieval Models

Dense: (7 models)

- **MiniLM** (all-MiniLM-L6-v2):
 - Lightweight general model (23M parameters)
- **MPNet** (all-mpnet-base-v2):
 - Lightweight general model (110M parameters)
- **SciNCL** (malteos/scincl):
 - Lightweight scientific model (110M parameters)
- **Specter** (allenai/specter):
 - Lightweight scientific model (125M parameters)
- **E5-Large** (intfloat/e5-large-v2):
 - Large general model (335M parameters)
- **GritLM-7B** (GritLM/GritLM-7B):
 - Large general model (7B parameters)
- **GTR-XL** (gtr-t5-xl):
 - Large general model (11B parameters)

Sparse: (1 model)

- **BM25:**
 - Keyword-based retrieval
 - Baseline

Hybrid: (7 models)

- Equally weighted aggregation of sparse and dense retrieval ($\alpha = 0.5$)
$$S_{\text{hybrid}}(d) = \alpha \cdot S_{\text{sparse}}(d) + (1 - \alpha) \cdot S_{\text{dense}}(d)$$
- Testing BM25 in combination with each dense retrieval model

Evaluation of Style Transfer (1/2)

		MRR@5	Baseline	Source Style: Original					Source Style: Formal				
				S0					S1				
Model		C0	C1	C2	C3	C4	C0	C1	C2	C3	C4		
BM25		0.5575	+0.0225	-0.0026	-0.0996	-0.0641	-0.0504	-0.0238	-0.0288	-0.0988	-0.0738		
MiniLM		0.4897	-0.0002	+0.0050	-0.0275	-0.0103	+0.0040	+0.0091	+0.0214	-0.0164	+0.0013		
MPNet		0.5052	+0.0094	+0.0133	+0.0099	+0.0091	-0.0113	-0.0026	+0.0126	+0.0028	-0.0022		
SciNCL		0.3735	+0.0031	+0.0058	-0.0095	-0.0178	-0.0210	-0.0186	+0.0034	-0.0177	-0.0280		
Specter		0.0728	+0.0067	+0.0137	+0.0544	-0.0046	+0.0406	+0.0535	+0.0727	+0.1004	+0.0462		
E5-Large		0.6568	+0.0010	-0.0011	-0.0456	-0.0251	-0.0523	-0.0419	-0.0273	-0.0824	-0.0556		
GritLM-7B		0.7115	-0.0025	-0.0066	-0.0466	-0.0428	-0.0256	-0.0320	-0.0355	-0.0776	-0.0517		
GTR-XL		0.5322	-0.0054	-0.0209	-0.0661	-0.0233	-0.0017	-0.0026	-0.0163	-0.0653	-0.0388		
Hybrid Retrieval Models	H-MiniLM	0.6248	+0.0122	-0.0010	-0.0852	-0.0524	-0.0362	-0.0193	-0.0161	-0.0825	-0.0597		
	H-MPNet	0.6323	+0.0134	+0.0021	-0.0722	-0.0472	-0.0388	-0.0173	-0.0105	-0.0758	-0.0563		
	H-SciNCL	0.6027	+0.0155	+0.0066	-0.0926	-0.0601	-0.0395	-0.0233	-0.0213	-0.0938	-0.0656		
	H-Specter	0.5523	+0.0122	-0.0162	-0.1471	-0.0674	-0.0306	-0.0025	-0.0113	-0.0994	-0.0518		
	H-E5-Large	0.6642	+0.0085	-0.0019	-0.0523	-0.0551	-0.0435	-0.0303	-0.0264	-0.0767	-0.0712		
	H-GritLM-7B	0.6898	+0.0084	-0.0026	-0.0517	-0.0437	-0.0361	-0.0219	-0.0166	-0.0616	-0.0519		
	H-GTR-XL	0.6370	+0.0087	-0.0113	-0.1086	-0.0600	-0.0362	-0.0286	-0.0320	-0.1030	-0.0723		

C1: Formal

C2: Scientific

C3: Abstract

C4: Question

C1: Formal
C2: Scientific
C3: Abstract
C4: Question

Evaluation of Style Transfer (2/2)

MRR@5		Baseline	Source Style: Summary					Source Style: Tweet				
		S0	S2					S3				
Model		C0	C0	C1	C2	C3	C4	C0	C1	C2	C3	C4
BM25		0.5575	-0.1034	-0.0776	-0.0905	-0.1790	-0.1184	-0.3086	-0.2902	-0.3206	-0.4177	-0.3450
MiniLM		0.4897	+0.0073	+0.0110	+0.0180	-0.0109	-0.0005	-0.0470	-0.0488	-0.0477	-0.0799	-0.0613
MPNet		0.5052	-0.0174	-0.0058	+0.0079	-0.0074	+0.0016	-0.0485	-0.0502	-0.0527	-0.0876	-0.0701
SciNCL		0.3735	-0.0159	-0.0106	+0.0022	-0.0320	-0.0158	-0.0536	-0.0574	-0.0564	-0.0682	-0.0570
Specter		0.0728	+0.0496	+0.0625	+0.0733	+0.0763	+0.0592	+0.0860	+0.0961	+0.1142	+0.0884	+0.0933
E5-Large		0.6568	-0.0399	-0.0266	-0.0264	-0.0775	-0.0433	-0.1035	-0.0985	-0.1032	-0.1205	-0.1444
GritLM-7B		0.7115	-0.0371	-0.0320	-0.0355	-0.0776	-0.0517	-0.1140	-0.0320	-0.0355	-0.0776	-0.0517
GTR-XL		0.5322	-0.0418	-0.0479	-0.0612	-0.1017	-0.0625	-0.0468	-0.0440	-0.0626	-0.0910	-0.0819
Hybrid Retrieval Models	H-MiniLM	0.6248	-0.0680	-0.0591	-0.0575	-0.1291	-0.0966	-0.2516	-0.2439	-0.2529	-0.2861	-0.2728
	H-MPNet	0.6323	-0.0728	-0.0628	-0.0623	-0.1264	-0.0985	-0.2628	-0.2530	-0.2558	-0.3042	-0.2721
	H-SciNCL	0.6027	-0.0790	-0.0589	-0.0747	-0.1588	-0.1078	-0.2677	-0.2577	-0.2719	-0.3233	-0.2943
	H-Specter	0.5523	-0.0701	-0.0503	-0.0659	-0.1739	-0.0959	-0.2465	-0.2284	-0.2400	-0.2935	-0.2652
	H-E5-Large	0.6642	-0.0814	-0.0681	-0.0747	-0.1300	-0.1064	-0.2760	-0.2601	-0.2644	-0.2892	-0.2892
H-GritLM-7B		0.6898	-0.0628	-0.0578	-0.0544	-0.1141	-0.0979	-0.2648	-0.2151	-0.2118	-0.2346	-0.2456
H-GTR-XL		0.6370	-0.0770	-0.0729	-0.0909	-0.1702	-0.1145	-0.2497	-0.2424	-0.2558	-0.2859	-0.2792

C1: Formal
C2: Scientific
C3: Abstract
C4: Question

Evaluation of ChecktThat! 2025 Task 4b

- **Submission:**
 - Based on the previous results we **submitted GritLM-7B without style transfer**
- **Findings:**
 - **Rank 12** in evaluation phase (Test)
 - **Test dataset** is more **challenging:**
 - BM25: -0.16 (for Test)
 - GritLM-7B: -0.12 (for Test)

Model	Dev	Test
BM25 (<i>Baseline</i>)	0.56	0.43
GritLM-7B (<i>Ours</i>)	<u>0.71</u>	<u>0.59</u>

Lessons Learned

GritLM-7B without style transfer achieves best retrieval performance:
MRR@5 = 0.71 (Dev) & 0.59 (Test)

Sparse vs. Dense vs. Hybrid: BM25 remains a strong sparse baseline, but dense E5-Large and GritLM-7B as well as all hybrid models generally outperform it

Transferring **style of claims** into a more **formal style boosts** retrieval (especially for BM25), while transferring **sources mostly decreases** performance

Style Combinations: Keeping the original sources, while using a formal or scientific style for claims gives the strongest results across most models



Thanks

Git Repository



Contact

Tobias Schreieder

TU Dresden & ScaDS.AI Dresden/Leipzig

Dresden, Germany

tobias.schreieder@tu-dresden.de



GEFÖRDERT VOM



Bundesministerium
für Bildung
und Forschung



SACHSEN Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.