

NumEval Task 3 @ SemEval 2024 - Numeral Aware Headline Generation

Hinoki Crum

School of Information, University of Arizona

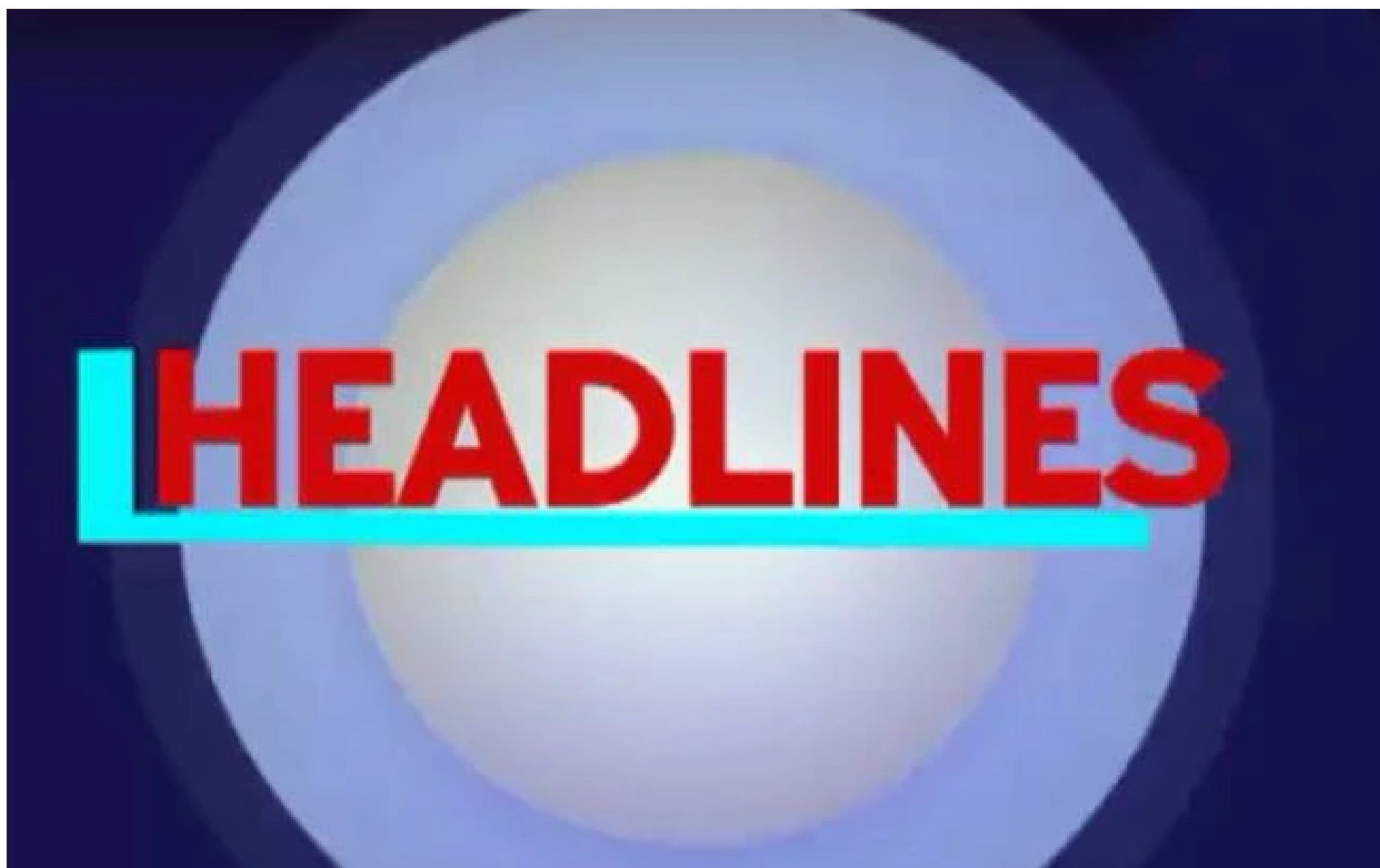


Objectives

Subtask 1: To compute the correct number to fill the blank in a news headline.

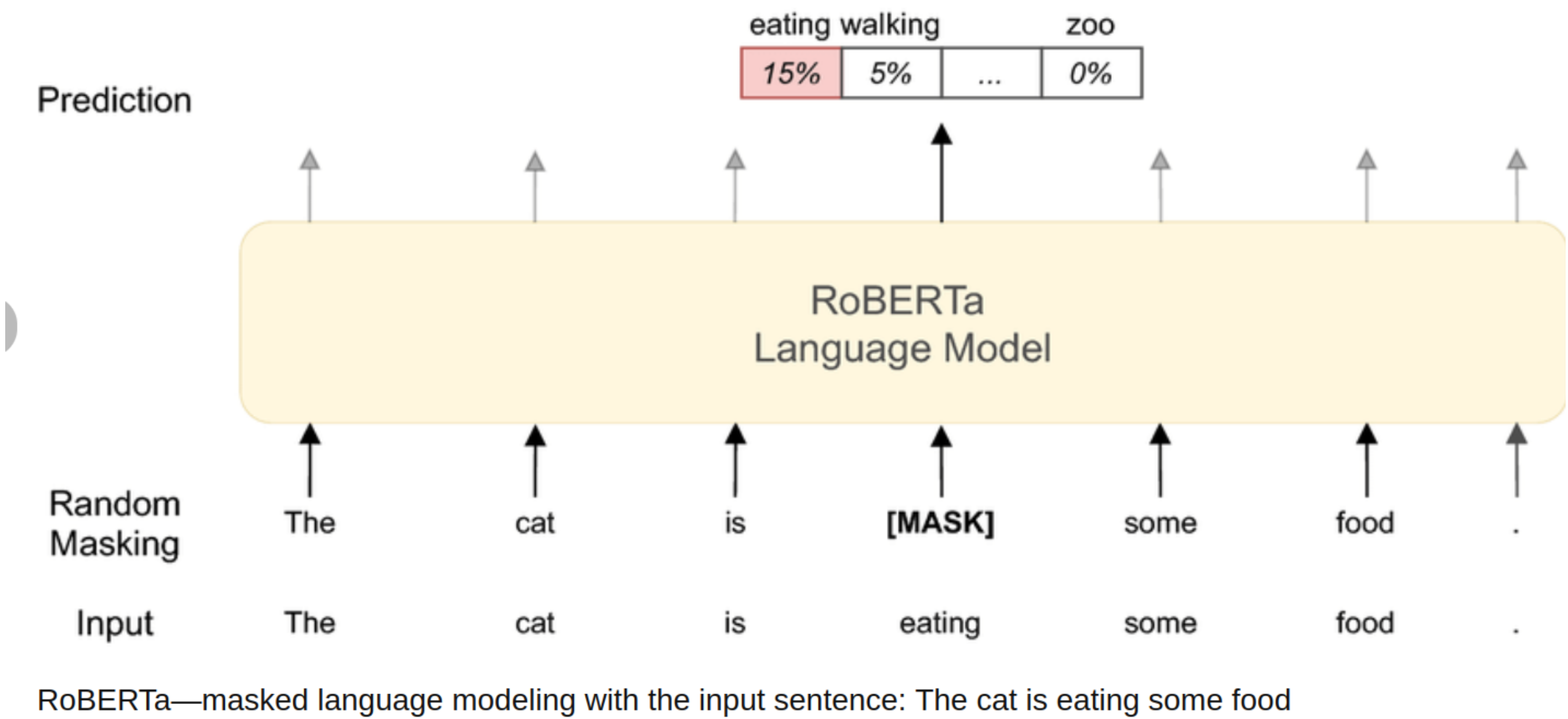
ans	calculation	masked headline	news
3	Trans(three)	___rd Victim Dead in Quarry Shooting; Manhunt...	(Oct 5, 2011 12:11 PM CDT) Police are still h...
12	Copy(12)	NYT Corrects 1853 Piece After ___ Years Win	(Mar 4, 2014 11:30 AM) The New York Times fol...
305	Round(305.45,0)	Stocks Up ___ in Election Rally	(Nov 4, 2008 3:19 PM) Stocks rallied on Elect...
1	Copy(1)	You Can Watch The Interview at ___pm	(Dec 24, 2014 11:19 AM) Turns out you won't e...
10	Copy(10)	Murder Suspect, ___, Will Stay in Adult Jail	(Oct 16, 2014 3:02 AM CDT) Tristen Kurilla, t...

Subtaks 2: To generate entire headlines on the provided news.

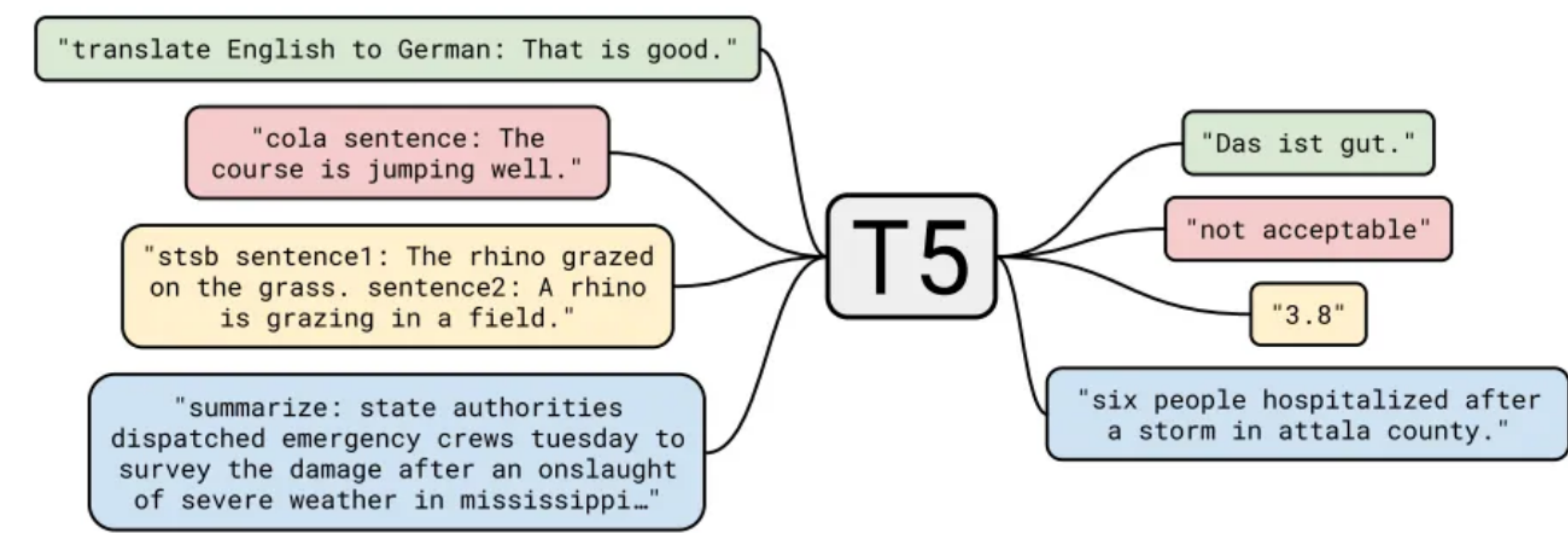


Models

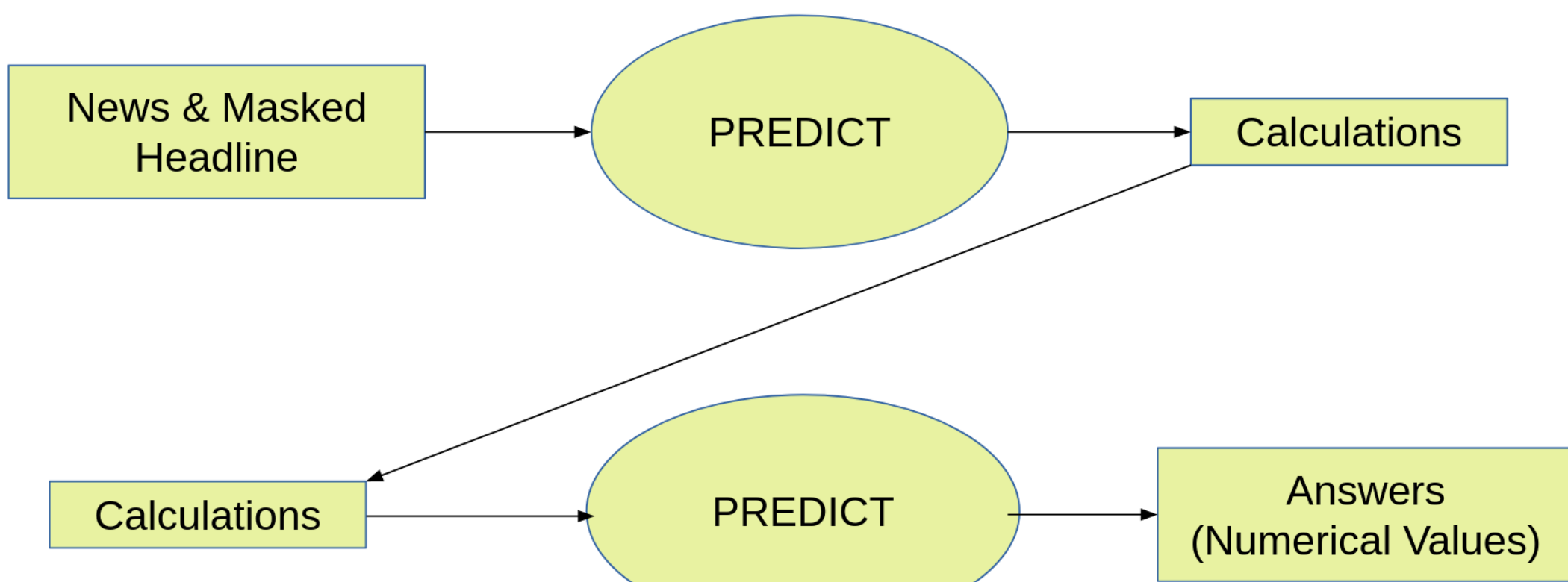
1. DistilRoBERTa (Masked Language Model)
2. T5 Language Model (Text-to-Text-Transfer-Transformer Model)
3. Flan-T5 (Enhanced Version of T5)



T5: Text-to-Text Transfer Transformer



2 Step Procedure for Subtask 1



Results

Table 1: Results for Subtask 1

Accuracy	One Step	Two Steps
DistilRoBERTa	79.8%	N/A
T5-Headline-Michal	87.7%	87.9%
T5-Title-Caleb	87.8%	88.1%
LaMini-Flan-T5	88.6%	90.2%

Table 2: Sample of Wrong Calculations Generated by T5 Model

Actual	Predicted
Round(Divide(268.30),0)	Copy(9)
Round(1.29,0)	Span(a trillion)
Subtract(Sept 5,July 8)	Subtract(30,7)
Add(22,Trans(four))	Add(Trans(four),22)
Multiply(Trans(one-quarter),100)	Multiply(Divide(Trans(one-quarter),100)

Mistakes are often made with the arithmetic operations, rounding the decimal numbers and the combination of various operations.

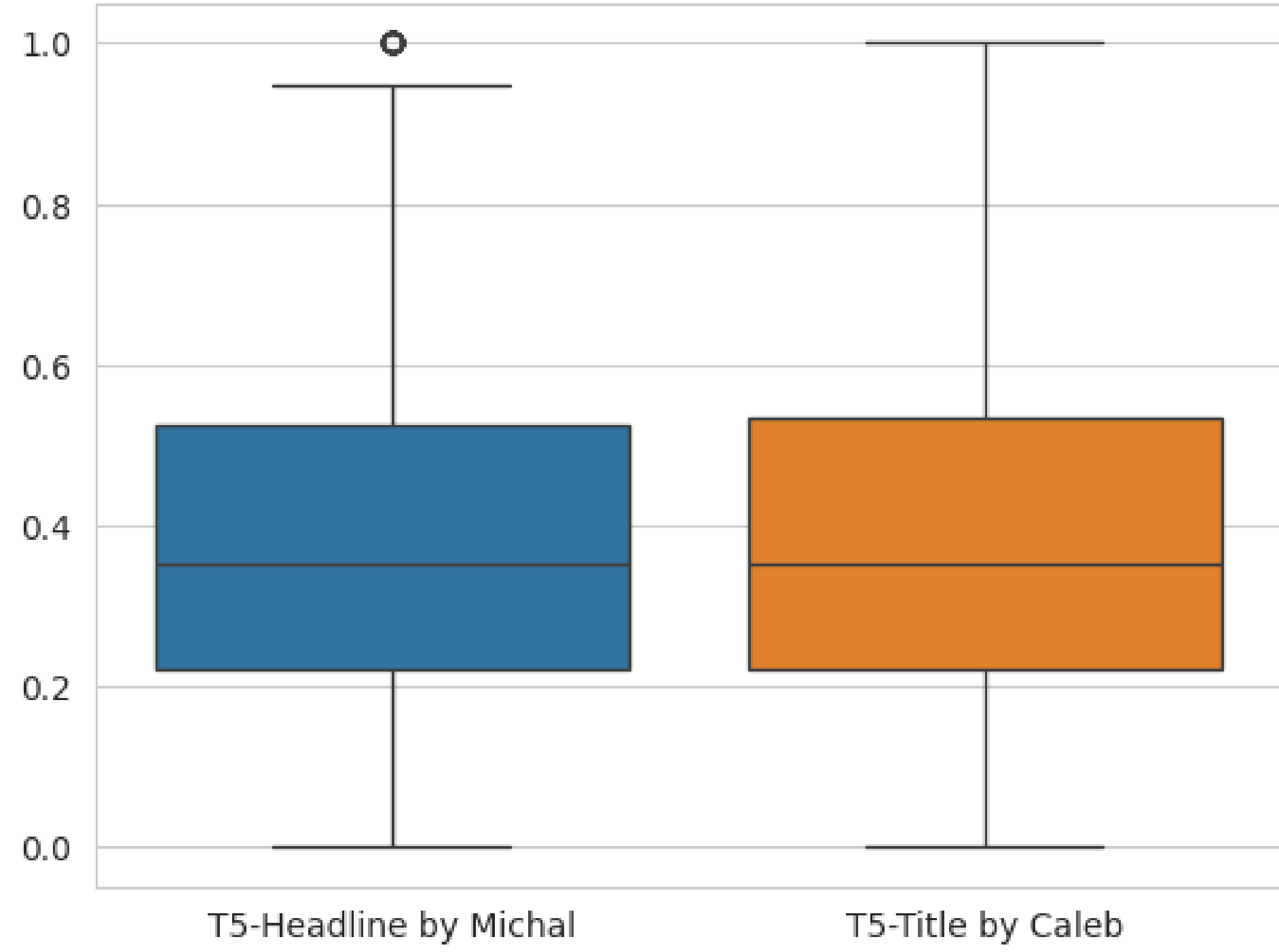


Figure 1: Rouge Scores for Headlines Generated by T5 Models for Subtask 2

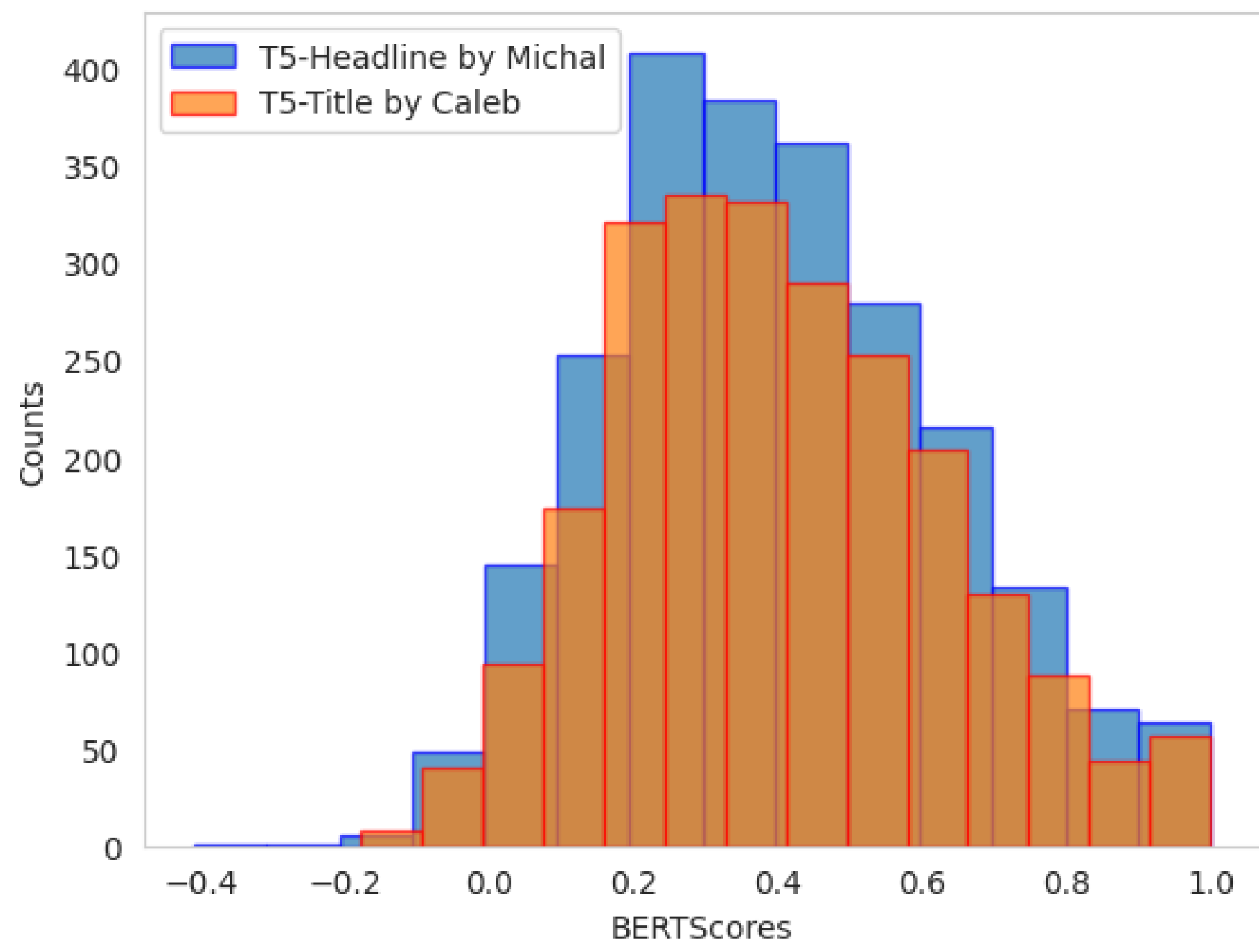


Figure 2: BERTScores for Headlines Generated by T5 Models for Subtask 2

Table 3: Table 3: Sample of Headline Generations by T5 Models

Type	Headline
Actual	25% of Freed Gitmo Detainees Returned to Terror
T5 Generated	1 in 4 Ex-Gatoramo Detainees Linked to Terrorists
Actual	She Offered \$25K to Catch a Killer. Now, She's Caught
T5 Generated	Woman Offers \$25K Reward in Husband's 2006 Murder
Actual	Ex-Congressman Caught With \$90K in Freezer Is Guilty
T5 Generated	Former Luisiana Rep Convicted of Stealing \$90K in Bribes
Actual	NBC Paid Chelsea Clinton \$600K a Year
T5 Generated	NBC Paid Chelsea Clinton \$600K a Year
Actual	Alec Baldwin Collects \$1.4K Every Time He Plays Trump
T5 Generated	Alec Baldwin's Trump Impersonation Is a SNL Sting
Actual	MH370 Decision Could Leave Mystery Unsolved
T5 Generated	3 Countries Suspended Flight 370 Search
Actual	We Drink and Drive an Estimated 121M Times a Year
T5 Generated	1.8% of US Adults Admit Driving While Impaired