

FAMuS: Frames Across Multiple Sources

Alexander Martin

Department of Computer Science, University of Rochester

Motivation

How can we improve knowledge distillation from large unstructured data?

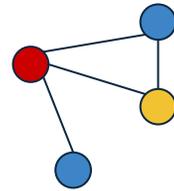
Motivation

How can we improve knowledge distillation from unstructured text?



Motivation

How can we improve knowledge distillation from unstructured text?



knowledge base



**Knowledge
distillation**

unstructured text



Template Extraction

On 16 July 2007 the inquiry was adjourned.

Template Extraction

On 16 July 2007 the inquiry was **adjourned**

Trigger Identification

Template Extraction

On 16 July 2007 the inquiry was **adjourned**

Activity_Pause		
Role	Definition	Filler
Activity	This FE identifies the Activity for which an Agent is pausing.	
Agent	An Agent pauses in the course of an Activity.	
Time	This FE identifies the Time at which an Agent pauses in the course of an Activity	
Place	This FE identifies the Place where an Agent pauses in the course of an Activity.	

Event Classification

Template Extraction

On 16 July 2007 the inquiry was adjourned

Activity_Pause		
Role	Definition	Filler
Activity	This FE identifies the Activity for which an Agent is pausing.	the inquiry
Agent	An Agent pauses in the course of an Activity.	
Time	This FE identifies the Time at which an Agent pauses in the course of an Activity	16 July 2007
Place	This FE identifies the Place where an Agent pauses in the course of an Activity.	

Argument Identification

Argument Classification

MegaWika: Report Passage



WIKIPEDIA
The Free Encyclopedia

Longdendale Bypass

Details of the public inquiry were announced in April 2007.^[26] It was to be run by Persona Associates with John Watson as Inspector.^[26] An official public inquiry website^[27] was launched offering news, transcripts and documentation for download.

The inquiry opened on 26 June 2007.^[26] On 16 July 2007 the inquiry was adjourned until 4 September with a final deadline for the submission of evidence of 14 August 2007.^[26] On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence,^[28] with further adjournments until 18 December.^{[26][29]}

MegaWika: Report Passage



WIKIPEDIA
The Free Encyclopedia

Longdendale Bypass

Details of the public inquiry were announced in April 2007.^[26] It was to be run by Persona Associates with John Watson as Inspector.^[26] An official public inquiry website^[27] was launched offering news, transcripts and documentation for download.

The inquiry opened on 26 June 2007.^[26] On 16 July 2007 the inquiry was **adjourned** until 4 September with a final deadline for the submission of evidence of 14 August 2007.^[26] On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence,^[28] with further adjournments until 18 December.^{[26][29]}

MegaWika: Source Document

Manchester
Evening News



News ▶ Greater Manchester News

Inquiry into bypass halted

THE public inquiry into the controversial Mottram-Tintwistle bypass was dramatically halted when the Highways Agency admitted it had got its figures wrong. All the traffic evidence it has put before the inquiry, which has been running since June, will now have to be reviewed.

FAMuS Overview

Report

On 16 July 2007 the inquiry was adjourned until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

FAMuS Overview

Report

On 16 July 2007 the inquiry was adjourned until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

FAMuS Overview

Report

On 16 July 2007 the inquiry was **adjourned** until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

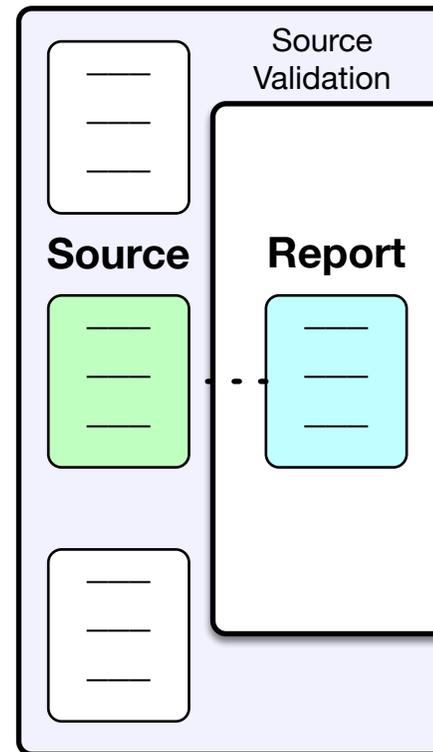
Source

THE public inquiry into the controversial Mottram - Tintwistle bypass was dramatically **halted** when the Highways Agency admitted it had got its figures wrong . All the traffic evidence it has put before the inquiry , which has been running since June , will now have to be reviewed ...

The agency told the inquiry , at **Stalybridge Civic Hall** , it had made a ` significant error ' in its estimation of how much traffic would use the road by underestimating journey times on the A616 trunk road ...

Inspector **John Watson** , chairing the inquiry , adjourned the hearing ... and said the new evidence should be heard and parties who have already given evidence be allowed to return .

Is the **highlighted event** in Report mentioned in the **Source**?



FAMuS Overview

Report

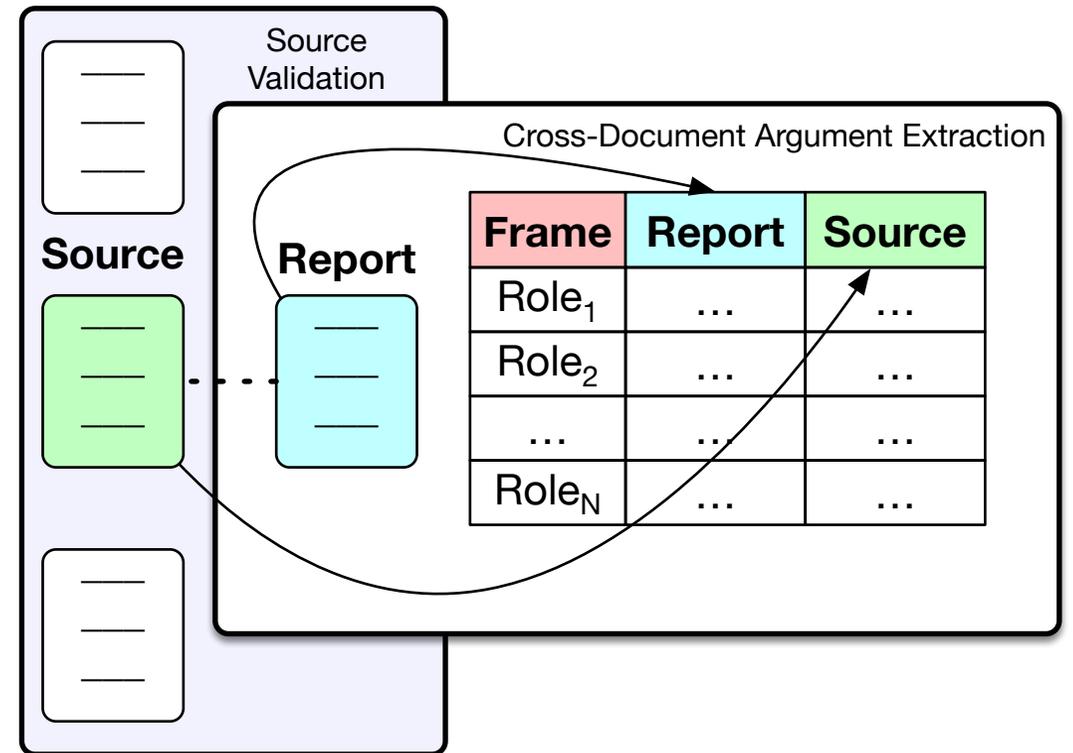
On 16 July 2007 the inquiry was **adjourned** until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

Source

THE public inquiry into the controversial Mottram - Tintwistle bypass was dramatically **halted** when the Highways Agency **admitted it had got its figures wrong** . All the traffic evidence it has put before the inquiry , which has been running since June , will now have to be reviewed ...

The agency told the inquiry , at **Stalybridge Civic Hall** , it had made a ` significant error ' in its estimation of how much traffic would use the road by underestimating journey times on the A616 trunk road ...

Inspector **John Watson** , chairing the inquiry , adjourned the hearing ... and said the new evidence should be heard and parties who have already given evidence be allowed to return .



FAMuS Overview

Report

On 16 July 2007 the inquiry was adjourned until 4 September with a final deadline for the submission of evidence of 14 August 2007 . On 11 September 2007 the inquiry was again adjourned until 19 September to allow the Highways Agency to review traffic evidence , with further adjournments until 18 December .

Source

THE public inquiry into the controversial Mottram - Tintwistle bypass was dramatically halted when the Highways Agency admitted it had got its figures wrong . All the traffic evidence it has put before the inquiry , which has been running since June , will now have to be reviewed ...

The agency told the inquiry , at Stalybridge Civic Hall , it had made a ' significant error ' in its estimation of how much traffic would use the road by underestimating journey times on the A616 trunk road ...

Inspector John Watson , chairing the inquiry , adjourned the hearing ... and said the new evidence should be heard and parties who have already given evidence be allowed to return .

Is the highlighted event in Report mentioned in the Source?

Source Validation

YES

Cross Document Argument Extraction

Event Type: Activity_pause

Roles	Report	Source
Activity	"the inquiry"	"THE public inquiry into the controversial Mottram - Tintwistle bypass"
Agent	-	"John Watson"
Place	-	"Stalybridge Civic Hall"
Time	"16 July 2007"	"when the Highways Agency admitted it had got its figures wrong"

FAMuS Data Collection

- Report-Source Pairs from MegaWika [1]

FAMuS Data Collection

- Report-Source Pairs from MegaWika [1]
- FrameNet ontology for Situations (Events, States, Processes)

FAMuS Data Collection

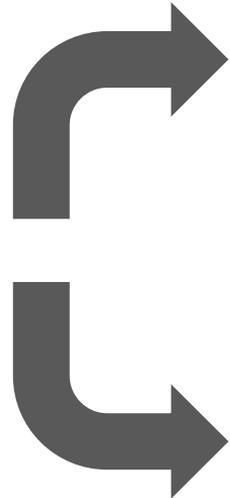
- Report-Source Pairs from MegaWika [1]
- FrameNet ontology for Situations (Events, States, Processes)

255 Situation Types

FAMuS Data Collection

- Report-Source Pairs from MegaWika [1]
- FrameNet ontology for Situations (Events, States, Processes)

255 Situations

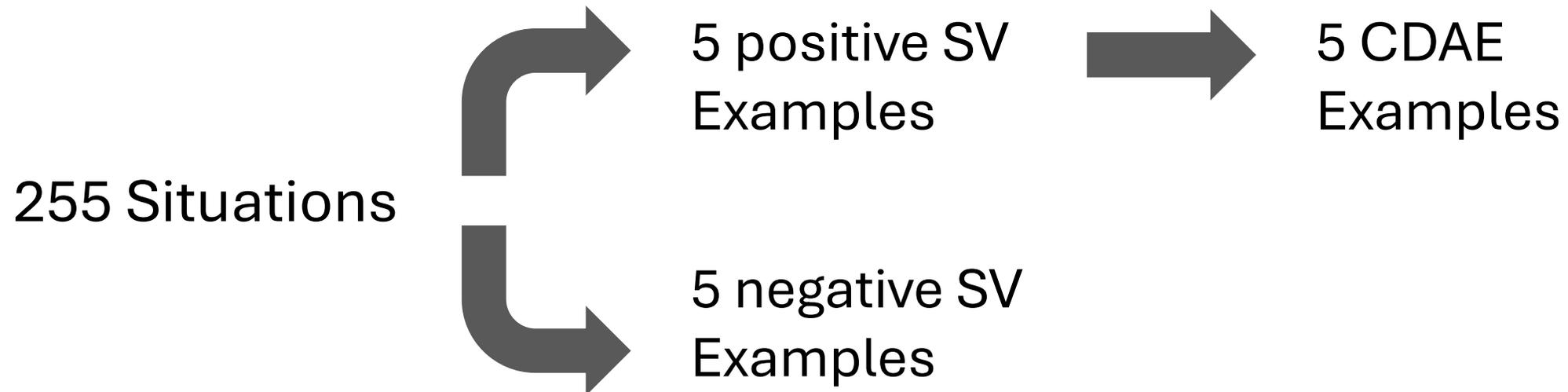


5 positive SV
Examples

5 negative SV
Examples

FAMuS Data Collection

- Report-Source Pairs from MegaWika [1]
- FrameNet ontology for Situations (Events, States, Processes)



Source Validation: Annotation

- Documents annotated by 3 MTURKers

Select Event Type:

None of the Event Type matches the highlighted span

2/3rd Agreement on event type

Does the Source Text contain the exact same event highlighted in the Passage Text?

Unanimous Agreement on Source Validation

CDAE: Annotation

- Annotate role fillers for each CORE role + Time and Place

Select Event Type:

Active Event Type:

Event Definition:
The words in this frame describe situations in which a Perpetrator carries off and holds the Victim against his or her will by force. " 'Two men kidnapped a Millwall soccer club employee, police said last night.'

Event Example: Two men **kidnapped** a Millwall soccer club employee, police said last night.

has been saved and has ... has invited the leaders of the world , including the Emperor Chin of Chinaland to a galactic peace conference . However the conference is disrupted when alien invaders capture **all members of the peace conference** and declare themselves rulers of the galaxy . To back up this claim , the invaders have assigned several champions as lieutenants . Hearing that things are once again in trouble , ninja warriors Ryu and Jack quickly enlist the help of the people of Futureland to build a spaceship and attack one of the champions .

Active Role: Answer:

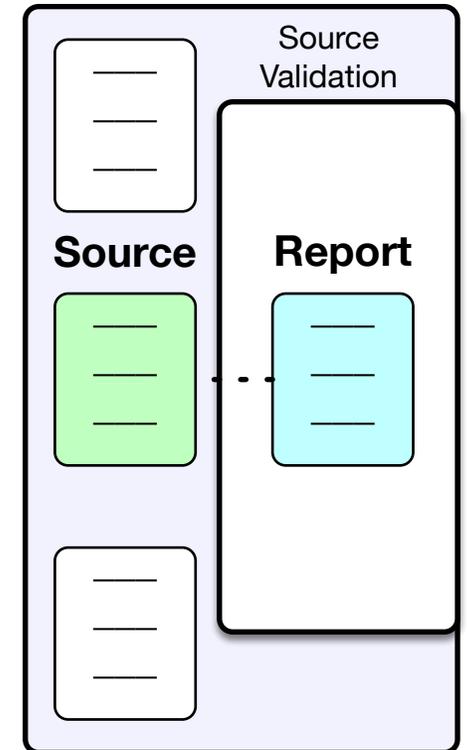
Roles:

Role Definition: The Victim is the person who is carried off and held against his/her will.

Is the **highlighted event** in **Report** mentioned in the **Source**?

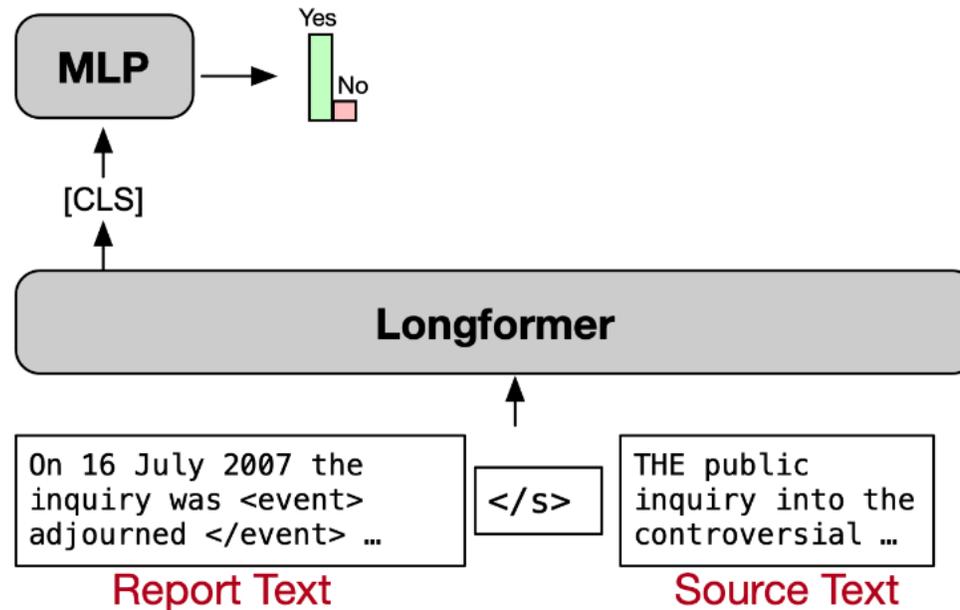
Source Validation: Models

- Lemma Baseline
 - Valid Source if report trigger is found in the source

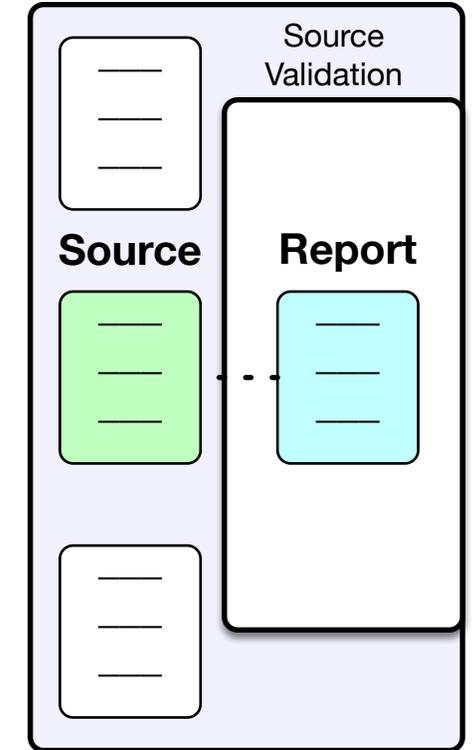


Source Validation: Models

- Lemma Baseline
 - Valid Source if report trigger is found in the source
- Finetuned Longformer Model



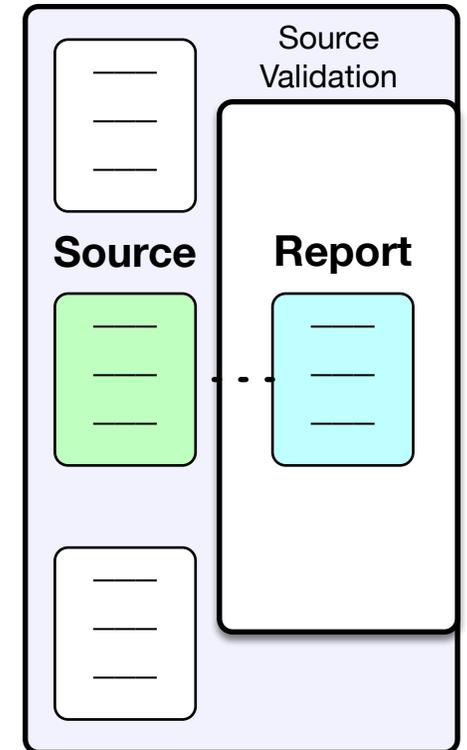
Is the **highlighted event** in **Report** mentioned in the **Source**?



Is the **highlighted event** in **Report** mentioned in the **Source**?

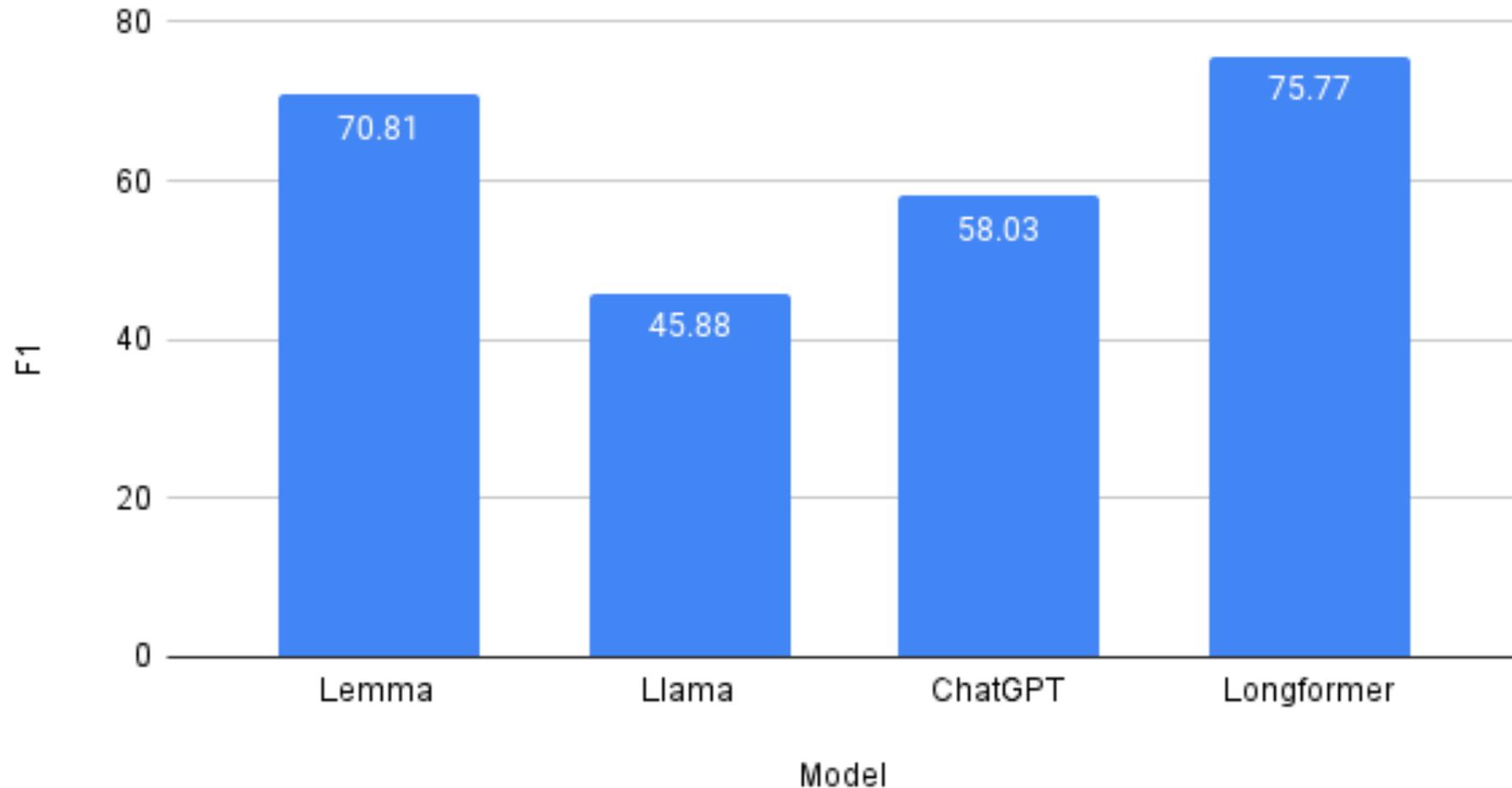
Source Validation: Models

- Lemma Baseline
 - Valid Source if report trigger is found in the source
- Finetuned Longformer Model
- LLM (4-shot incontext learning)

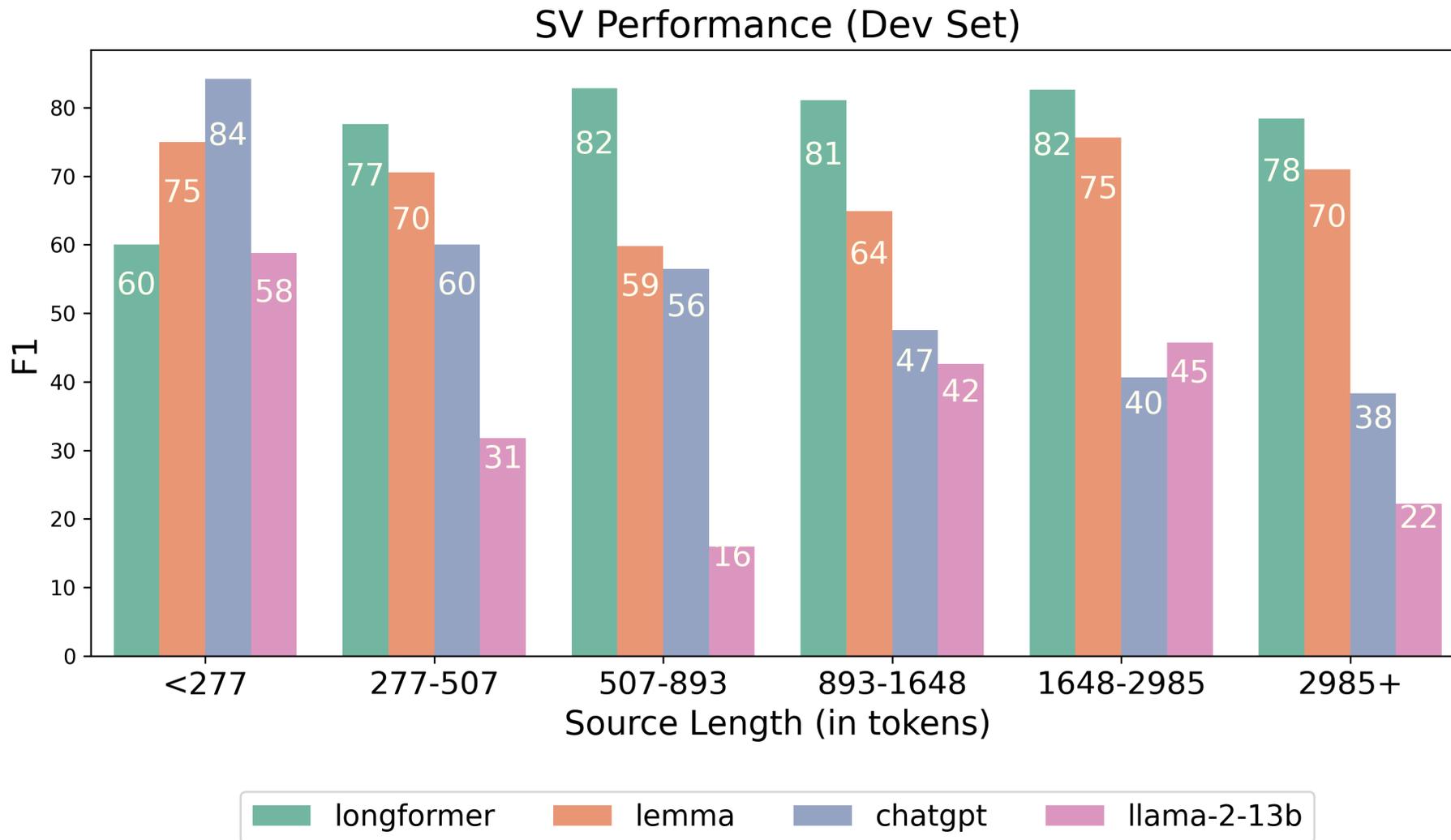


Source Validation: Results

F1 Performance of Source Validation Models



Document Length: Source Validation Results



CDAE: Settings

Inputs:

Report:

- Report Text (w/ tagged event)
- Event Type



Outputs:

Role fillers for the **Report Text**

CDAE: Settings

Inputs:

Report:

- Report Text (w/ tagged event)
- Event Type



Outputs:

Role fillers for the **Report Text**

Source:

- Report Text (w/ tagged event)
- Source Text
- Event Type



Role fillers for the **Source Text**
for the tagged report event

CDAE: QA

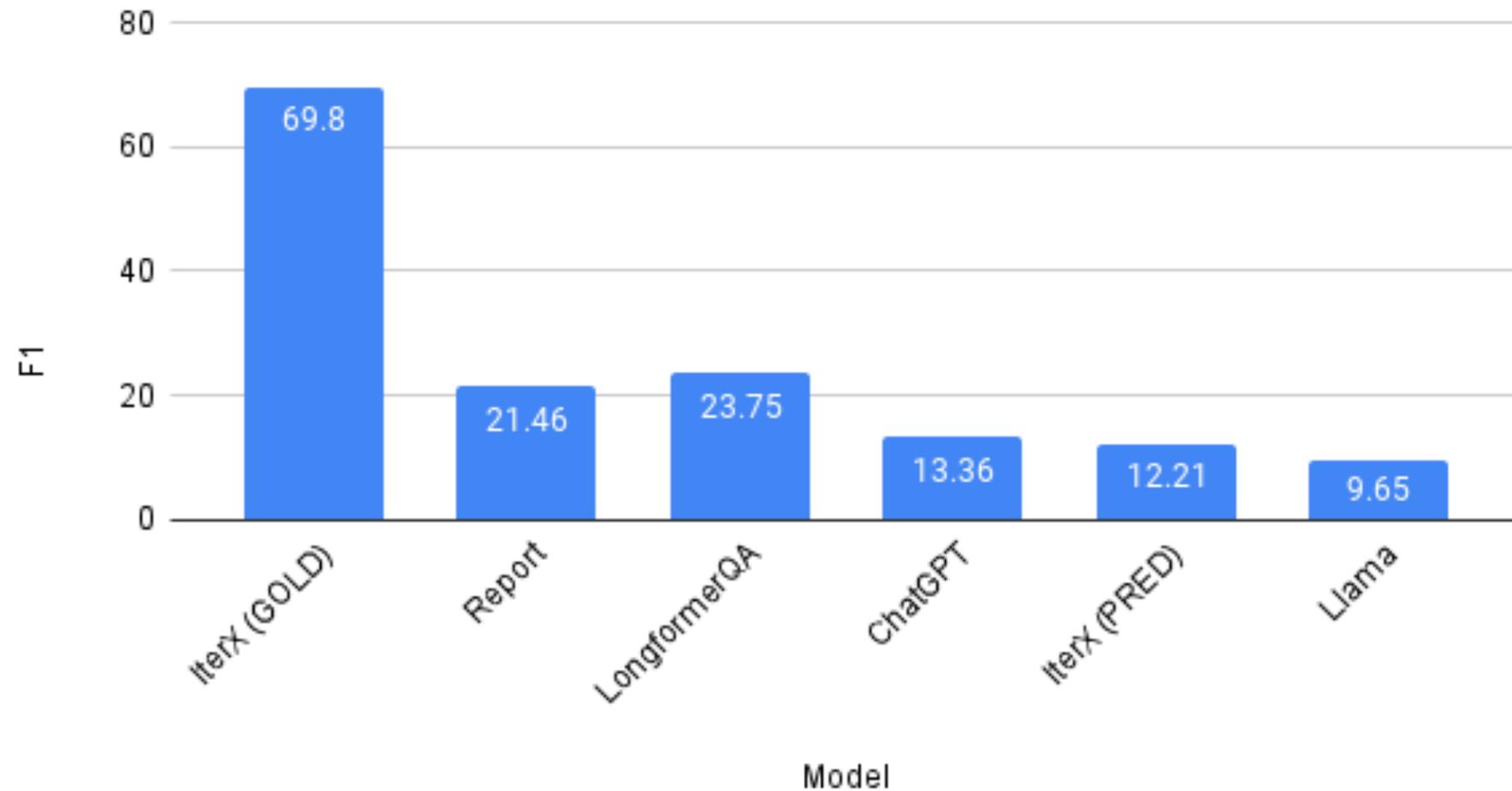
	Report Model	Source Model
Question	Event: Activity_pause, Role: Activity, Trigger: adjourned	Event: Activity_pause, Role: Activity, Report: On 16 July 2007 the inquiry was <code><event> adjourned </event></code> until 4 September with a ...
Context	On 16 July 2007 the inquiry was adjourned until 4 September with a ...	THE public inquiry into the controversial Mottram - Tintwistle bypass was dramatically halted when the Highways Agency admitted it had got its figures wrong ...
Answer	the inquiry	THE public inquiry into the controversial Mottram - Tintwistle bypass

CDAE: IterX

	Report Model	Source Model
Candidate Spans	<ol style="list-style-type: none">1. "<event> adjourned </event>"2. "16 July 2007",3. "the inquiry",4. "until 4 September" ...	<ol style="list-style-type: none">1. "<event> adjourned </event>"2. "THE public inquiry into the controversial Mottram - Tintwistle bypass"3. "public",4. "into the controversial Mottram - Tintwistle bypass" ...
Context	<event> adjourned </event> On 16 July 2007 the inquiry was adjourned until ...	On 16 July 2007 the inquiry was <event> adjourned </event> until 4 September ... THE public inquiry into the controversial Mottram - Tintwistle bypass was ...

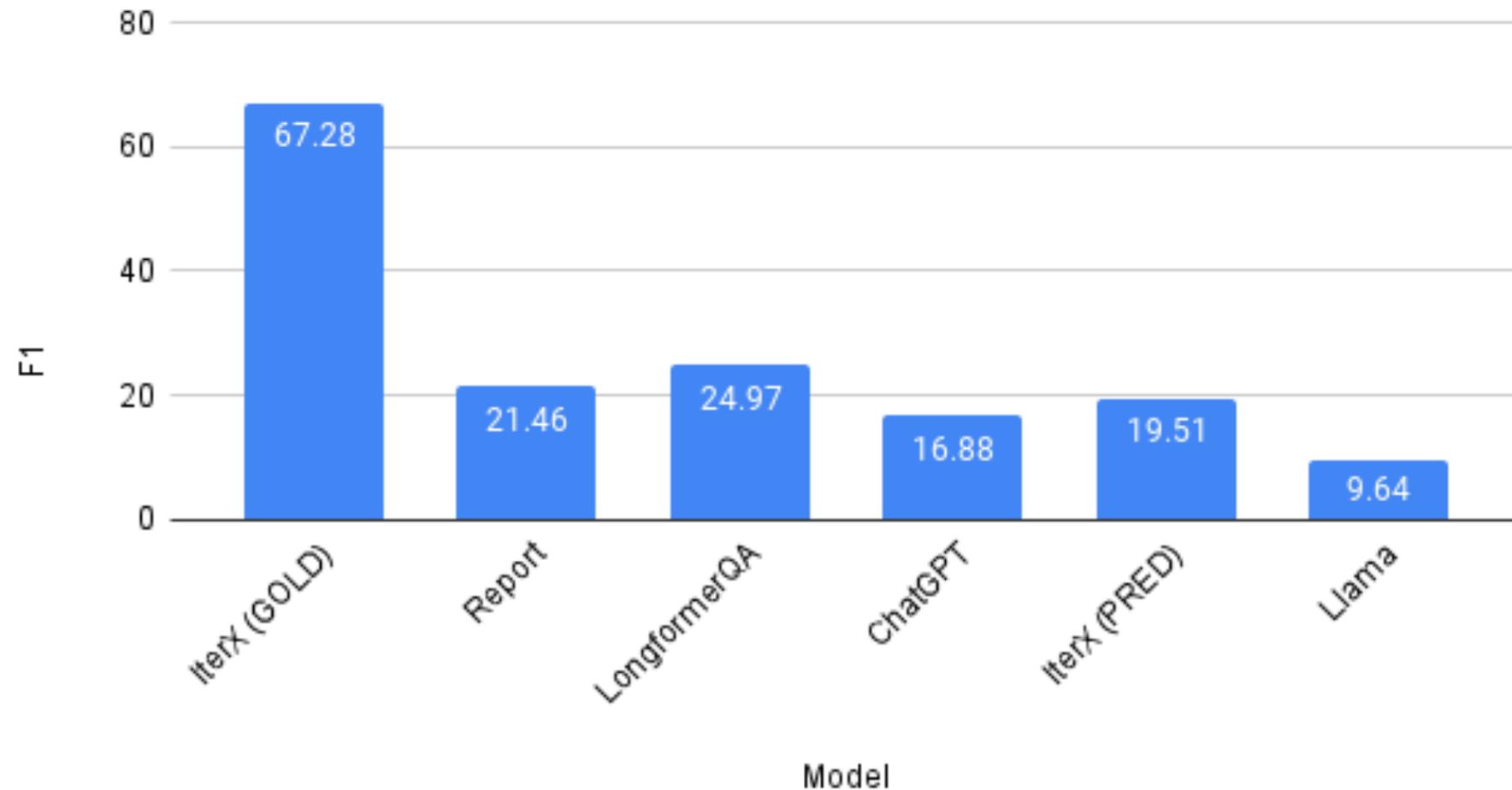
CDAE: Results

CEAF-RME F1 Performance of CDAE Models



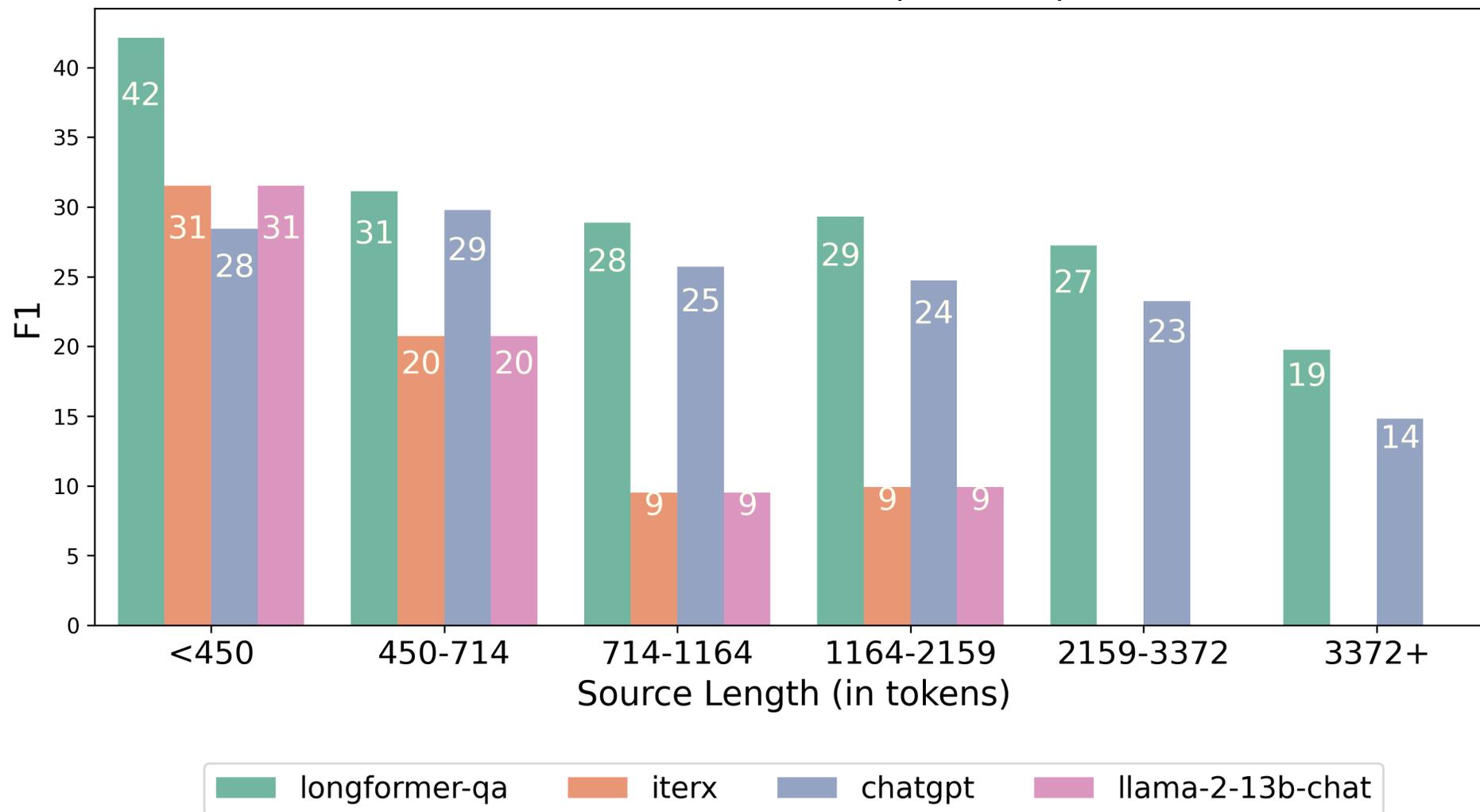
CDAE: Results + Report Baseline

CEAF-RME F1 Combined With Report Baseline



Document Length: CDAE Results

CDAE Performance (Dev Set)



FAMuS: Frames Across Multiple Sources

amart50@u.rochester.edu

<https://github.com/FACTSlab/FAMuS>



Sid Vashishtha
Rochester (now Google)



Will Gantt
Rochester



Ben Van Durme
JHU



Aaron Steven White
Rochester

Appendix

Source Validation: Negative Examples

Does the Source Text contain the exact same event highlighted in the Passage Text?

Yes	No
-----	----

Unanimous Agreement on Negative Source Validation

- Expert annotator qualifies the source as a negative example

HOWEVER,

We do not get 5 examples for each event type,
so we generate silver examples

Source Validation: Silver Negatives

- Frame f_i , has candidate pool of report-source pairs C_i

Source Validation: Silver Negatives

- Frame f_i , has candidate pool of report-source pairs C_i
- Sample report-source pair $(r_i^j, s_i^j) \in C_i$
- Sample another pair $(r_i^k, s_i^k) \in C_i - \{(r_i^j, s_i^j)\}$

Source Validation: Silver Negatives

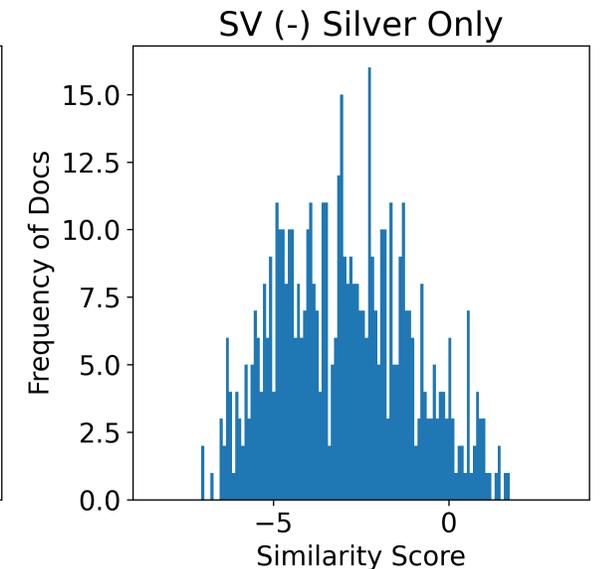
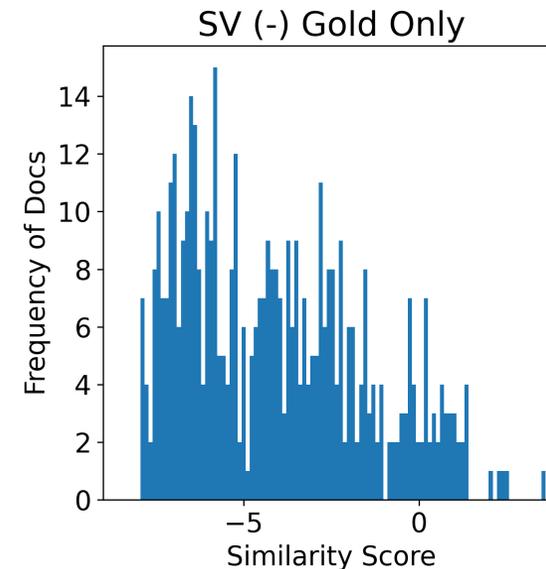
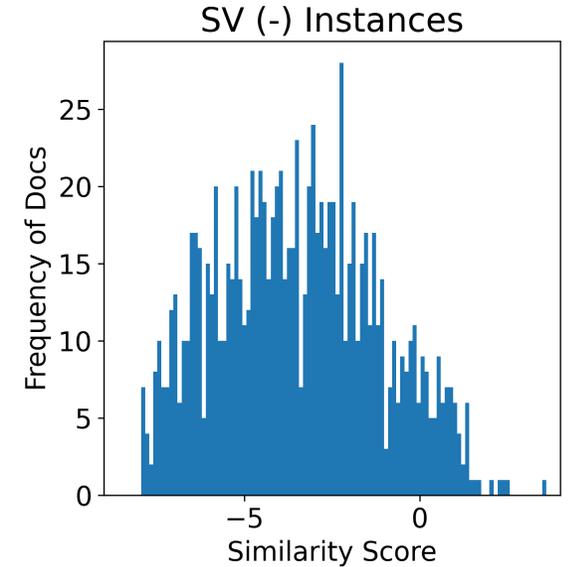
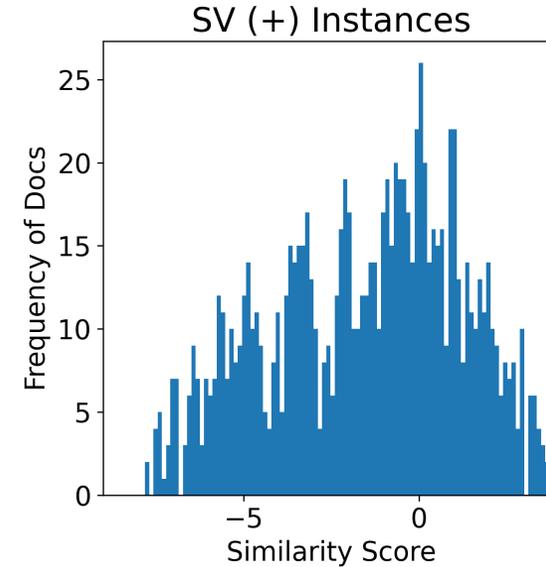
- Frame f_i , has candidate pool of report-source pairs C_i
- Sample report-source pair $(r_i^j, s_i^j) \in C_i$
- Sample another pair $(r_i^k, s_i^k) \in C_i - \{(r_i^j, s_i^j)\}$
- Negative example (r_i^j, s_i^k) , s.t. s_i^k has highest similarity to r_i^j
- Use SimLM [2] to compute similarity between r and s

What percent of SV are Negative?

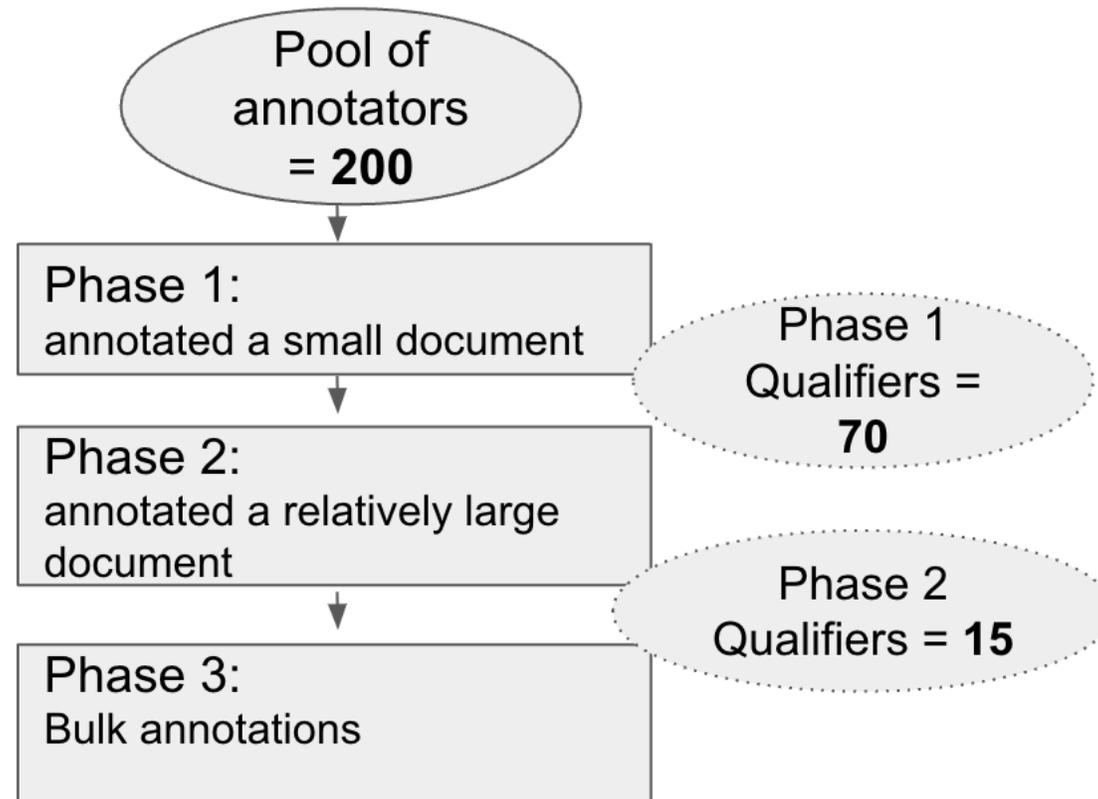
- 22% (556) of the 2250 total exams are silver

More on SV Instances

- Silver distribution matches annotated distribution in similarity well (slightly shifted towards 0)

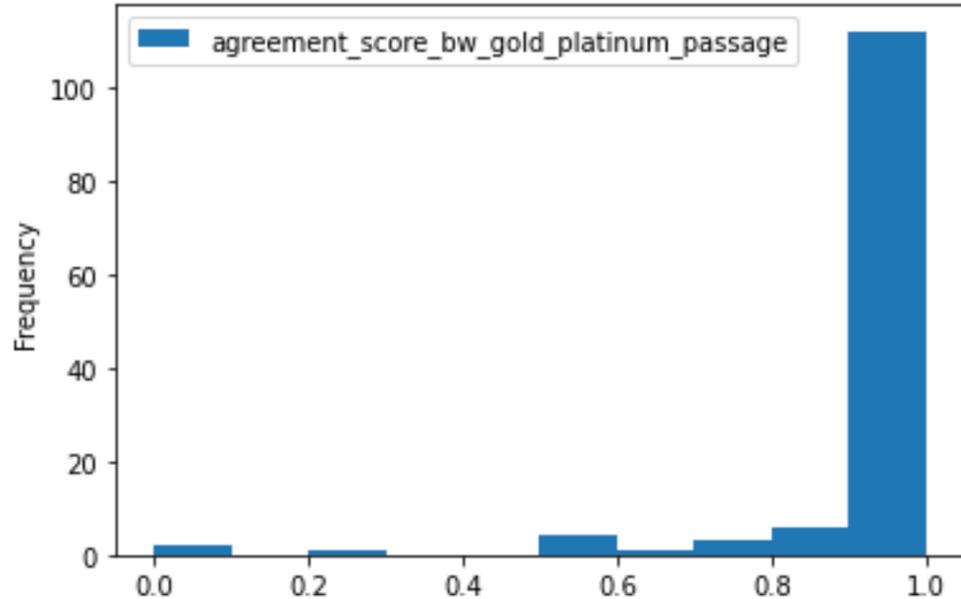


CDAE: Annotator Qualification

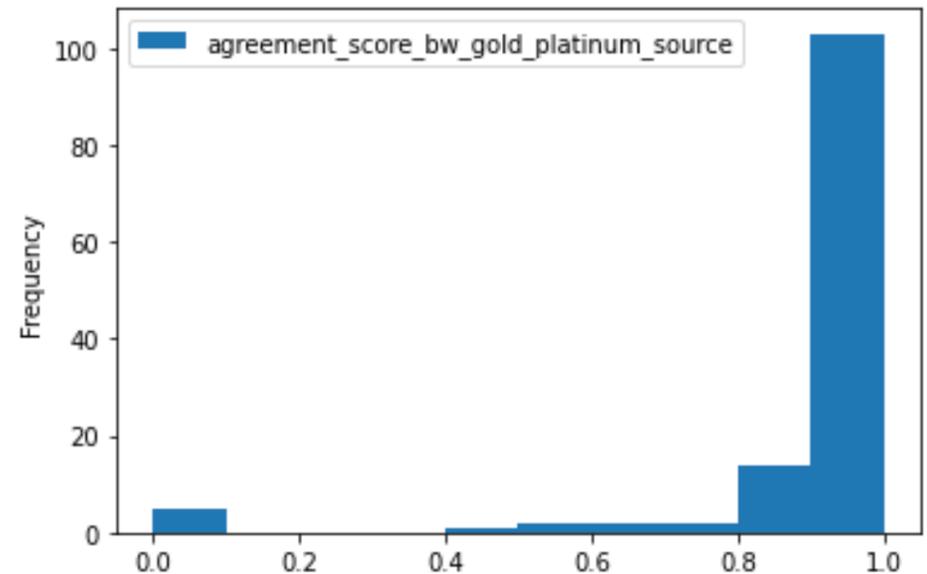


CDAE: Quality Assurance

- Launched annotations in iterations
- After iteration 1, expert annotators did platinum annotations on all 130 documents



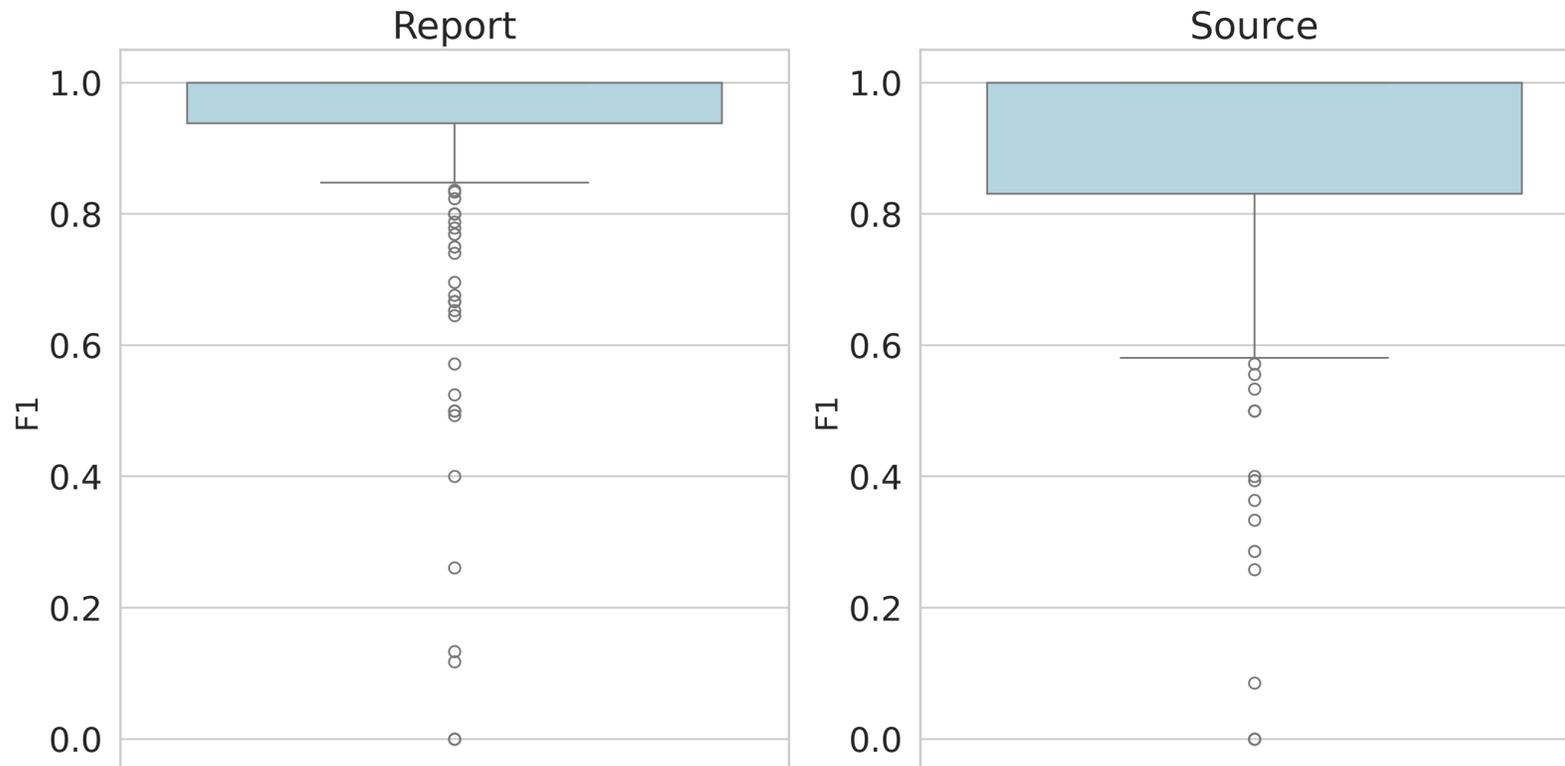
Report Roles Mean Agreement (F1): 94.42



Source Roles Mean Agreement (F1): 91.70

CDAE: Automatic Quality Assurance

- ChatGPT predicts roles and checked IAA with each annotator
- Experts annotate bottom 25% agreement from each batch



Cited

1. Barham et al. (2023). MegaWika: Millions of reports and their sources across 50 diverse languages.
2. Wang et al (2023). SimLM: Pre-training with Representation Bottleneck for Dense Passage Retrieval
- 3.