Controllable Text Generation

Controlling Style and Content

Shrimai Prabhumoye, 2022

NVIDIA

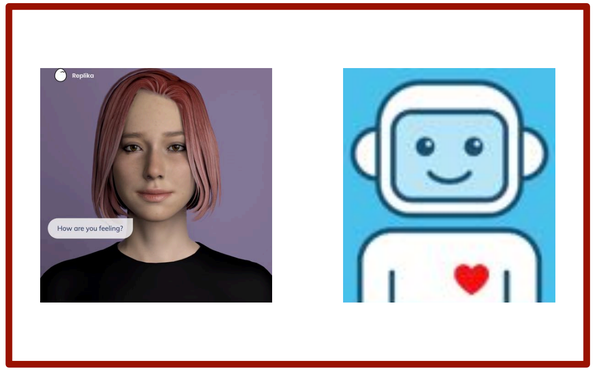
NLP Applications are on the rise!

NLP Applications are on the rise!

Personal Assistants

NLP Applications are on the rise!

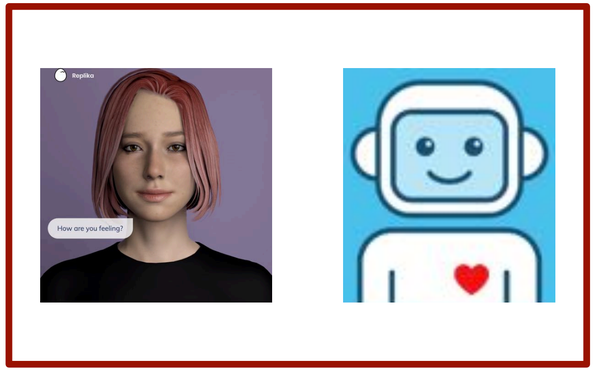
Personal Assistants



Empathy/Compassion bots

NLP Applications are on the rise!

Personal Assistants

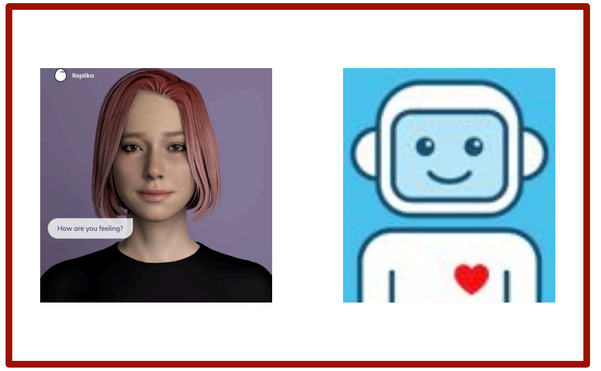


Empathy/Compassion bots

Call-center/FAQ bots

NLP Applications are on the rise!

Personal Assistants



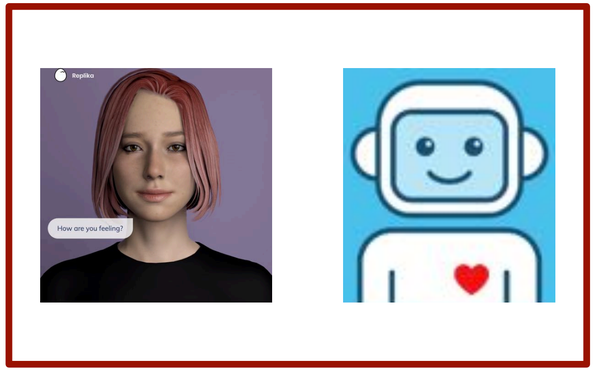
Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing

NLP Applications are on the rise!

Personal Assistants



Empathy/Compassion bots

Call-center/FAQ bots

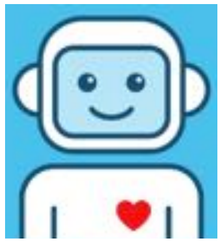
Ubiquitous Computing

Other Applications

NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

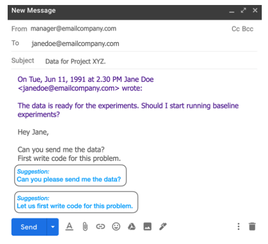
Personal Assistants



Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing



Other Applications

NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

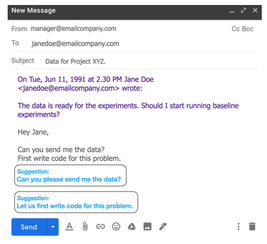
Personal Assistants

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing



Other Applications

NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

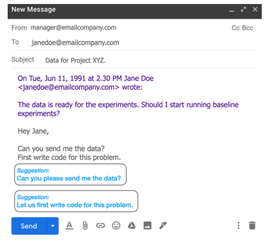
- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

- Content needs to be accurate!
- Structured/ Unstructured data

Call-center/FAQ bots

Ubiquitous Computing



Other Applications

NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Emotions (style)
- Sensitivity of content

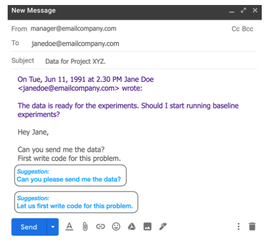
Empathy/Compassion bots

- Content needs to be accurate!
- Structured/ Unstructured data

Call-center/FAQ bots

- Accurate information
- Structured data

Ubiquitous Computing



Other Applications

NLP Applications are on the rise!

- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

- Emotions (style)
- Sensitivity of content

Empathy/Compassion bots

- Content needs to be accurate!
- Structured/ Unstructured data

Call-center/FAQ bots

- Accurate information
- Structured data

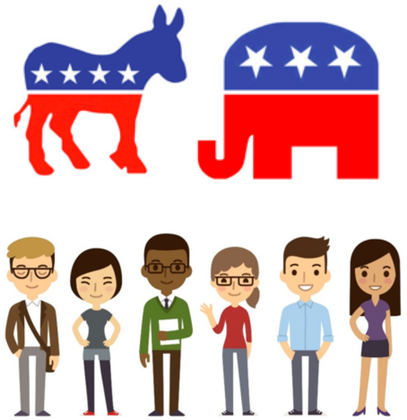
Ubiquitous Computing

- Plot of stories (structure)
- Sequence of topics in stories (structure)
- Politeness of emails (style)

Other Applications

Overview

Overview



Style

ACL 2018
ACL 2020

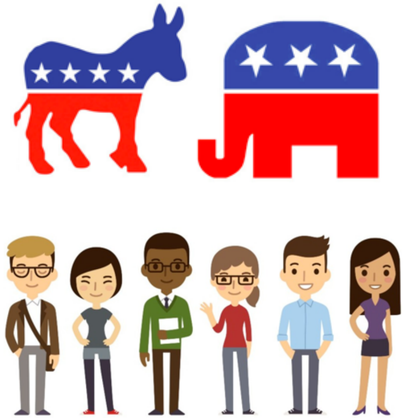
Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

ACL Findings
2022

Overview



Style

ACL 2018
ACL 2020

Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

ACL Findings
2022

What is Style Transfer?

- Rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context.

What is Style Transfer?

- Rephrasing the text to contain specific stylistic properties without changing the intent or affect within the context.

“Shut up! the video is starting!”

“Please be quiet, the video will begin shortly.”

Challenges in style transfer

Challenges in style transfer

- Style is *culturally diverse*

Usage of “please” to closest friends.

Challenges in style transfer

- Style is *culturally diverse*
- Style is *subtle*

Indirect: By the way, do you know when is the deadline?

1st person plural: Let us find a good name for this paper.

Challenges in style transfer

- Style is *culturally diverse*
- Style is *subtle*
- *Disentangle* style and content

Challenges in style transfer

- Style is *culturally diverse*
- Style is *subtle*
- *Disentangle* style and content
- *Lack* of parallel data

Goals

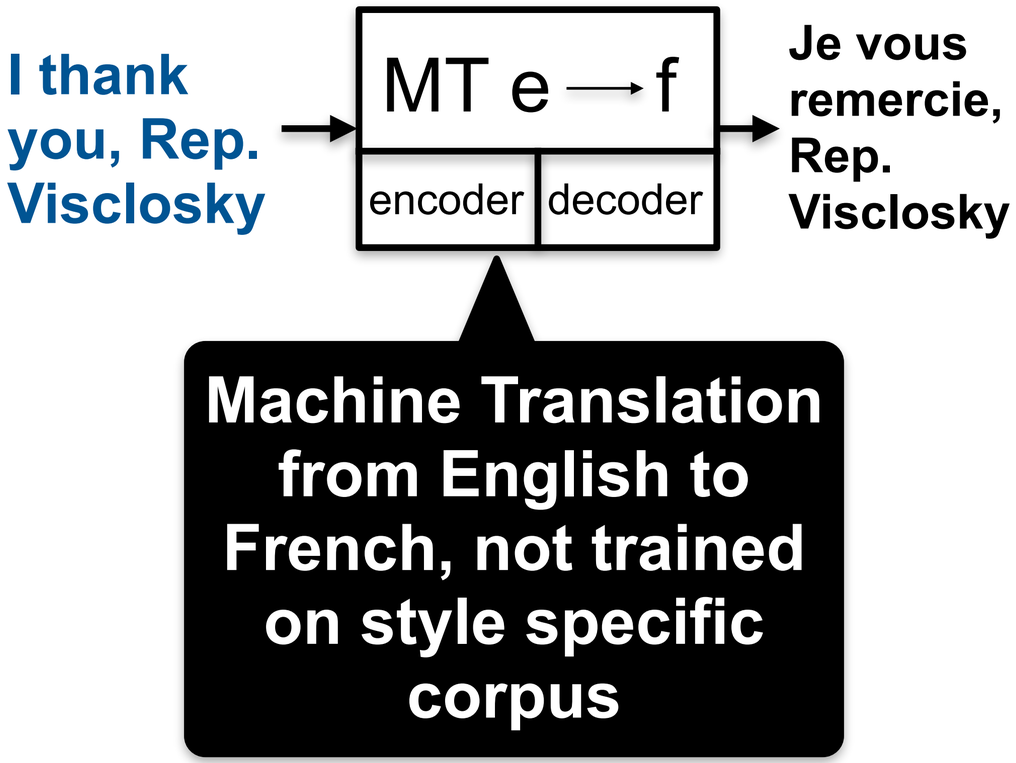
- ☐ Transfer the Style
- ☐ Preserve the meaning
- ☐ Representation that reduces stylistic properties
- ☐ Interpretable Representation

Back-Translation (BST)

(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation

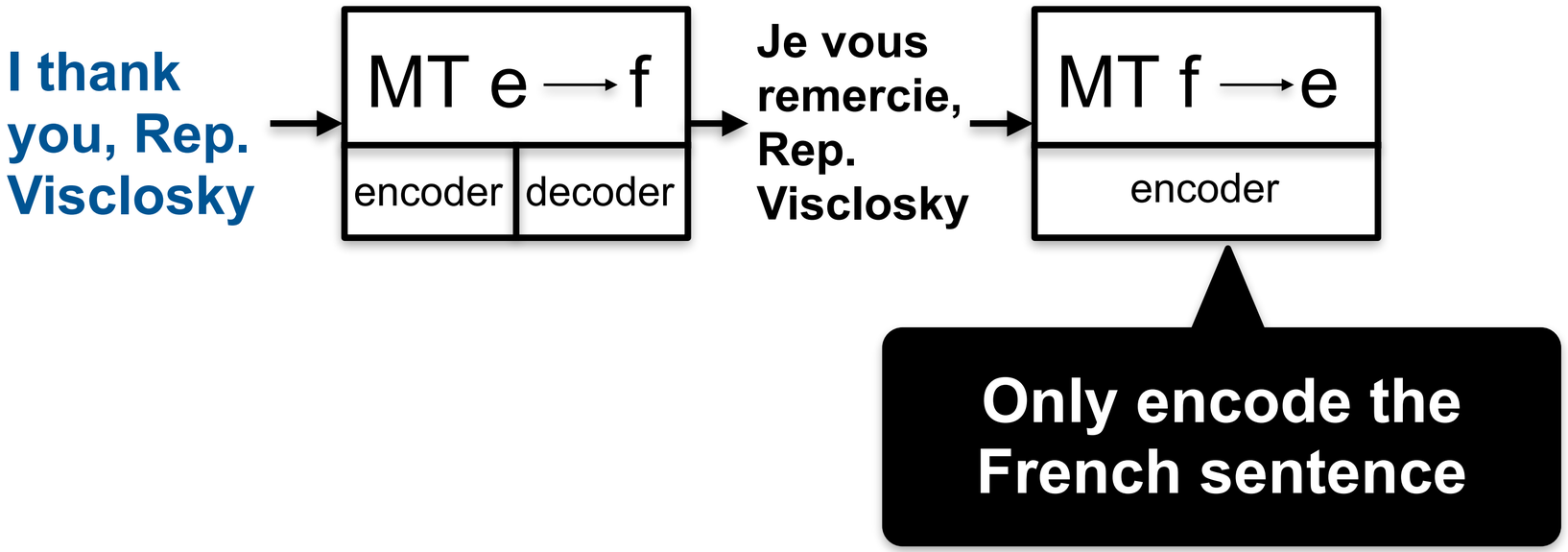
Back-Translation (BST)



(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation

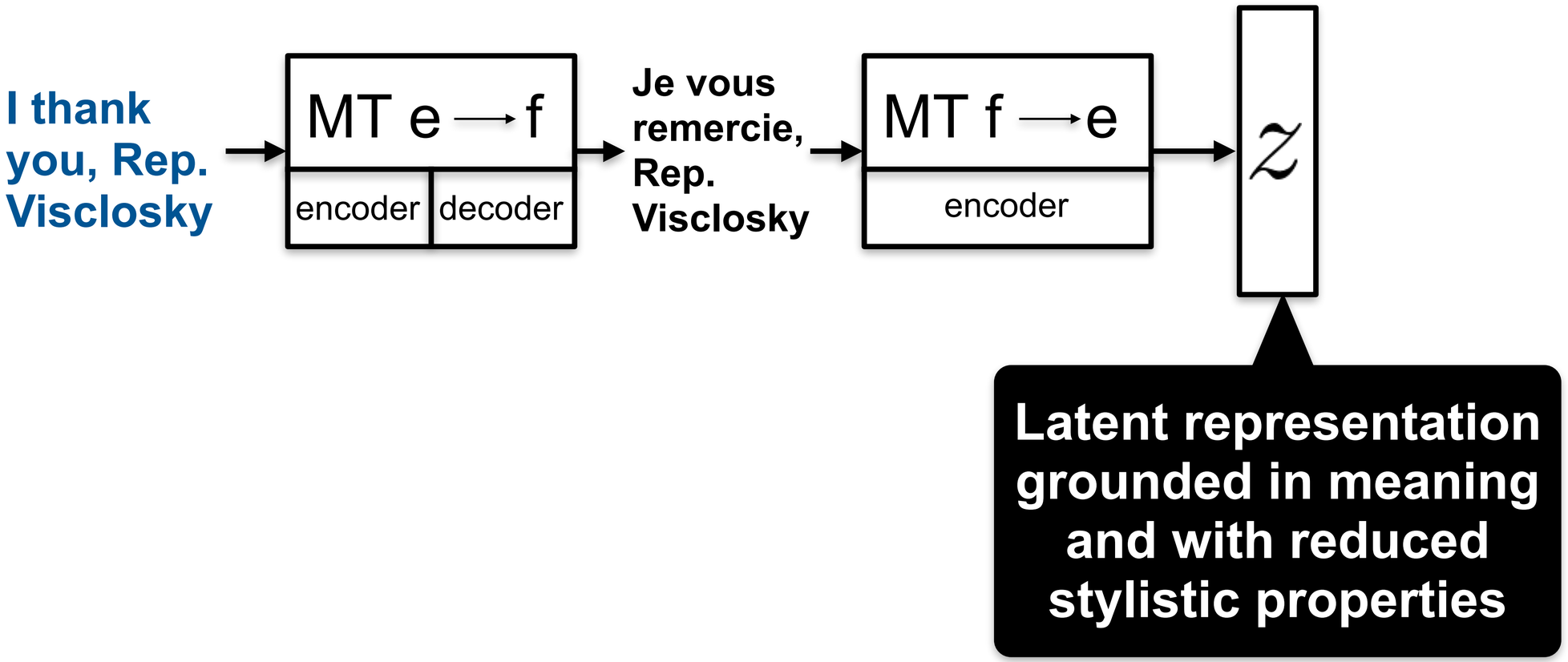
Back-Translation (BST)



(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation

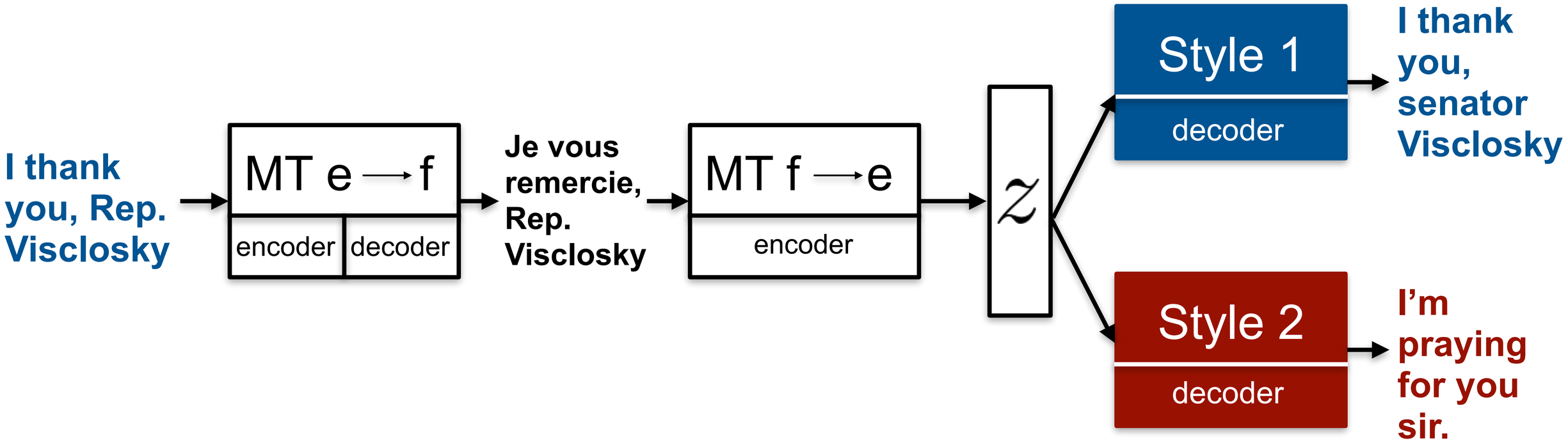
Back-Translation (BST)



(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation

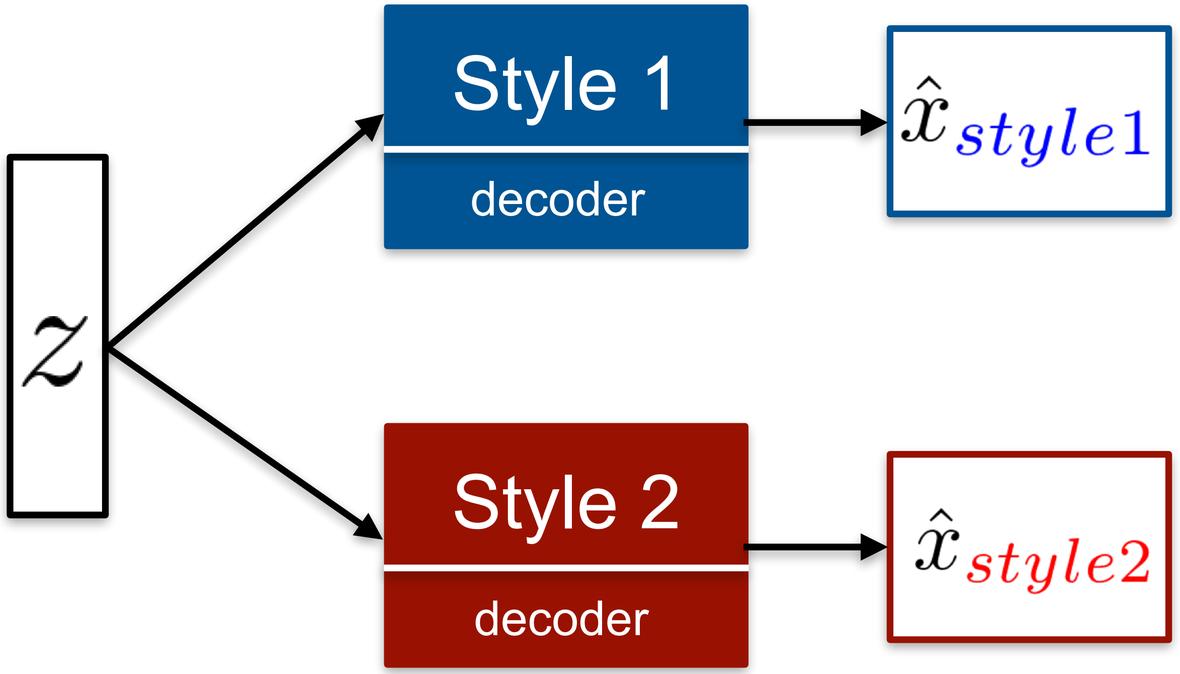
Back-Translation (BST)



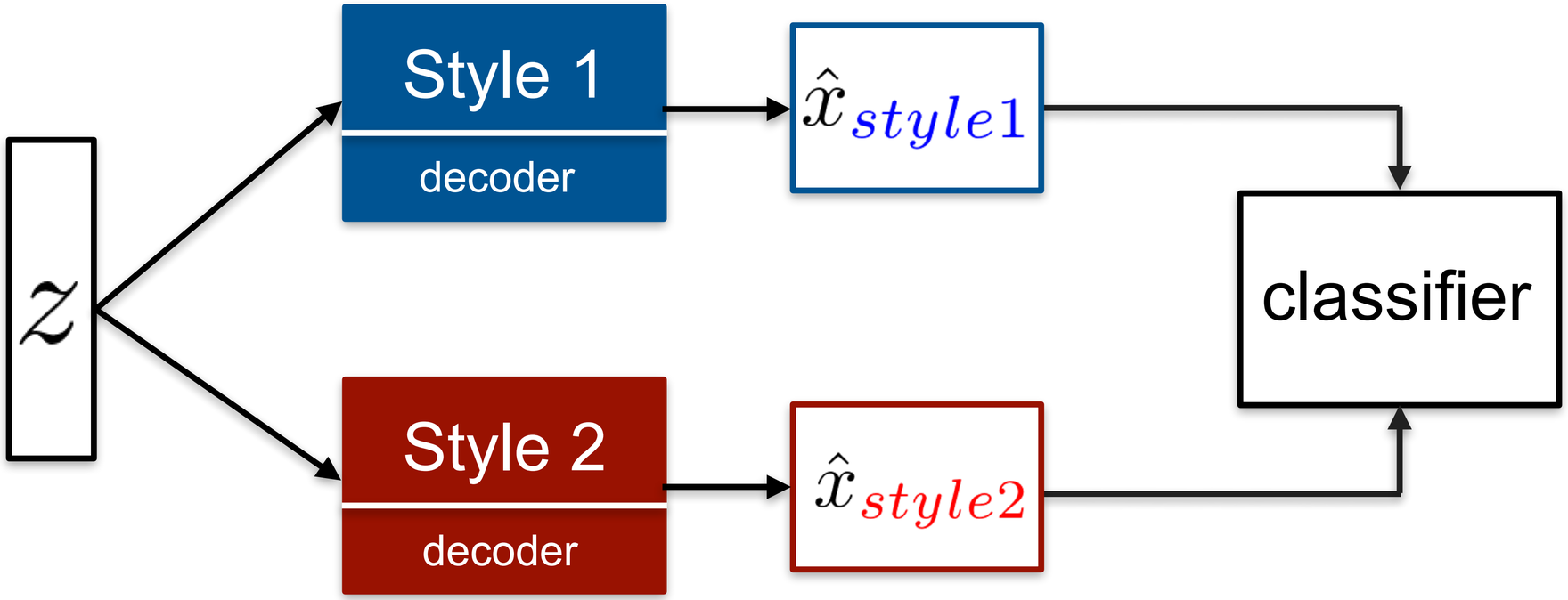
(Prabhumoye et al, ACL 2018)

https://github.com/shrimai/Style-Transfer-Through-Back-Translation

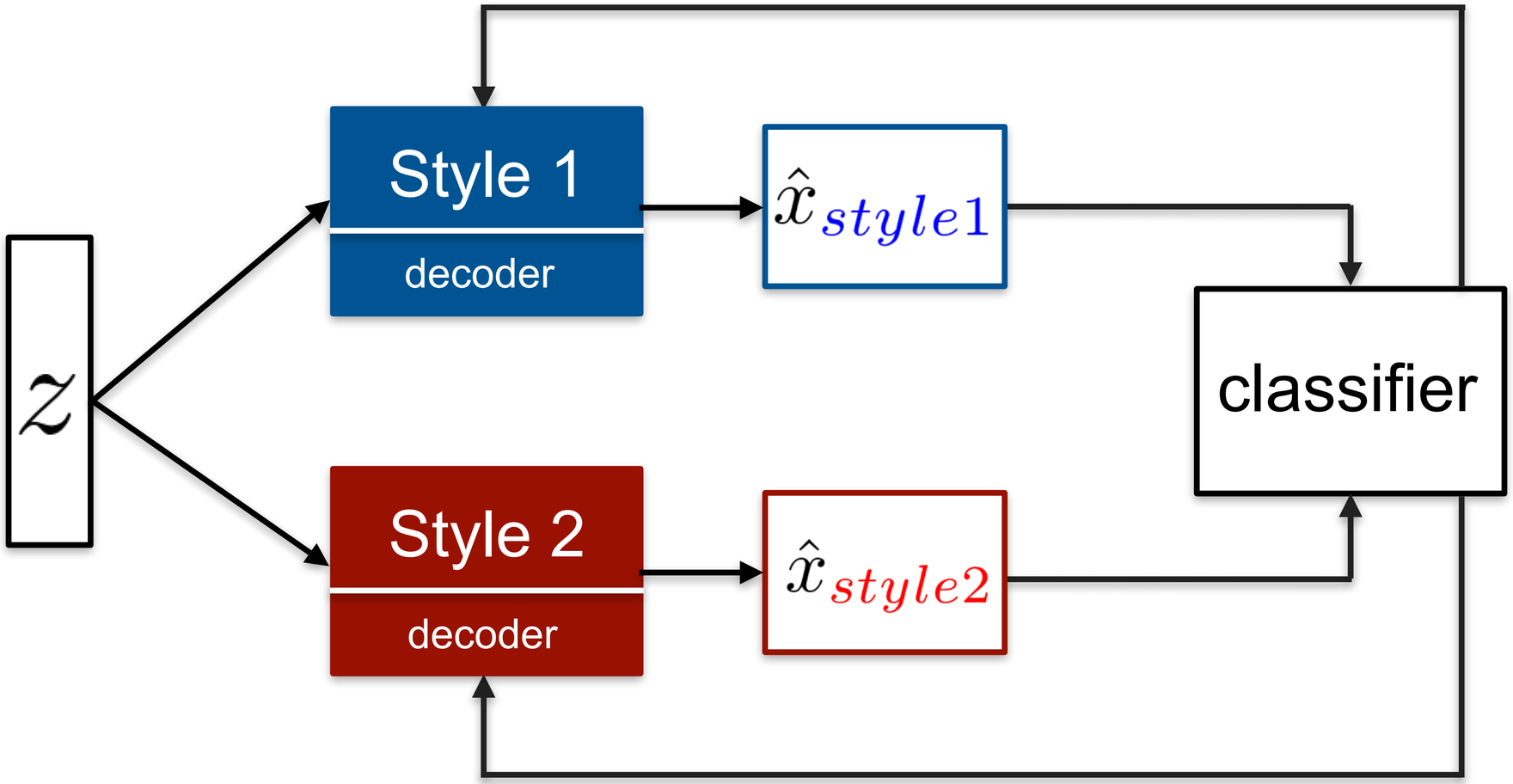
Training of Back-Translation



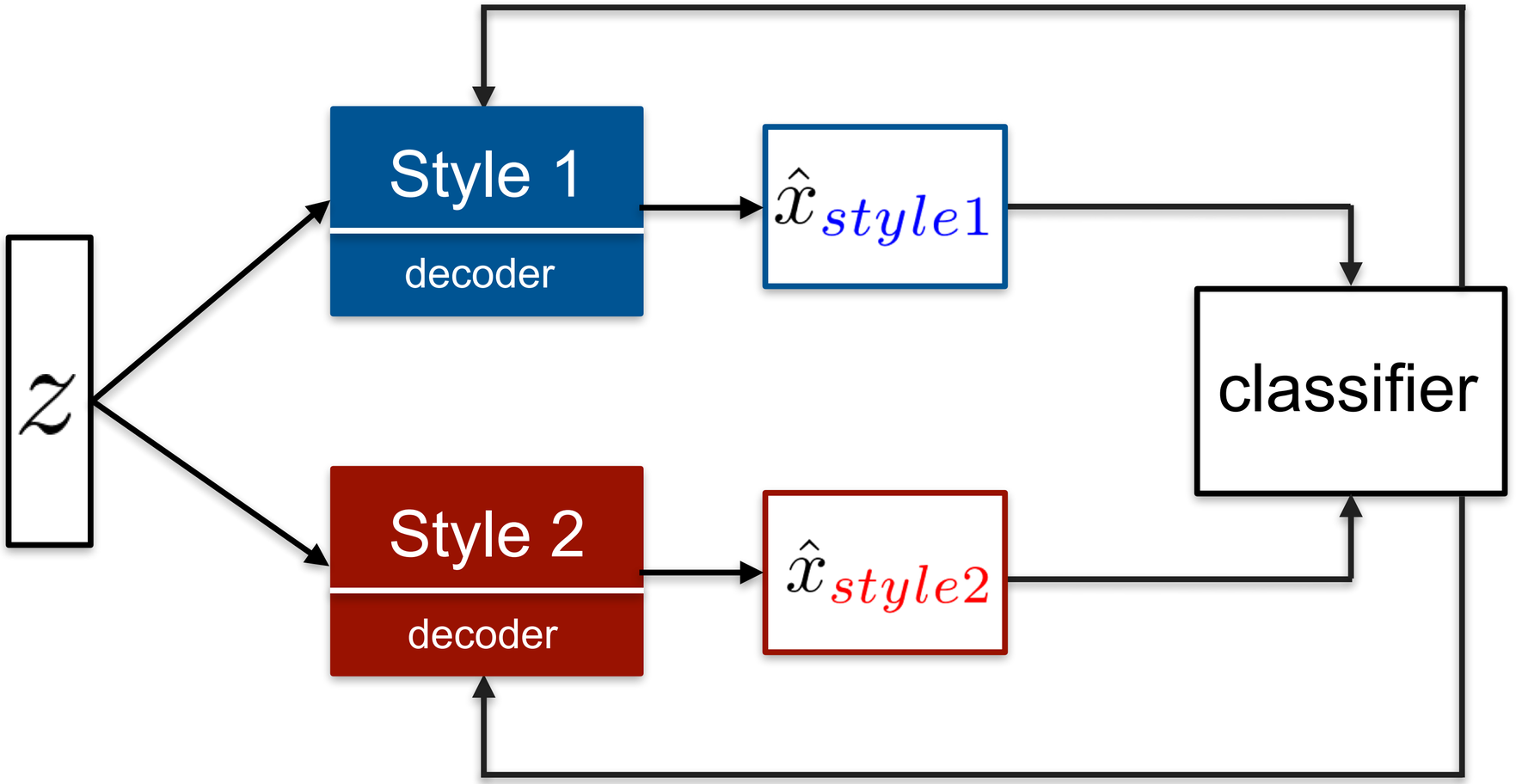
Training of Back-Translation



Training of Back-Translation



Training of Back-Translation



$$\min_{\theta_{\text{gen}}} \mathscr{L}_{\text{gen}} = \mathscr{L}_{\text{recon}} + \lambda_c \mathscr{L}_{\text{class}}$$

Goals achieved by Back-Translation

- ☐ Transfer the Style
- ☐ Preserve the meaning
- ☐ Representation that reduces stylistic properties
- ☐ Interpretable Representation

Goals achieved by Back-Translation

- [x] Transfer the Style

  feedback in the form of weighted classifier loss

- [ ] Preserve the meaning

- [ ] Representation that reduces stylistic properties

- [ ] Interpretable Representation

Goals achieved by Back-Translation

- [x] Transfer the Style

  feedback in the form of weighted classifier loss

- [x] Preserve the meaning

  styles are surface level, z is in latent space and LSTM

- [ ] Representation that reduces stylistic properties

- [ ] Interpretable Representation

Goals achieved by Back-Translation

- [x] Transfer the Style
  - feedback in the form of weighted classifier loss
- [x] Preserve the meaning
  - styles are surface level, z is in latent space and LSTM
- [x] Representation that reduces stylistic properties
  - translating to another language
- [ ] Interpretable Representation

Goals achieved by Back-Translation

- ☑ Transfer the Style
  - feedback in the form of weighted classifier loss
- ☒ Preserve the meaning
  - styles are surface level, z is in latent space and LSTM
- ☑ Representation that reduces stylistic properties
  - translating to another language
- ☒ Interpretable Representation
  - z is in latent space

Tag and Generate

(Madaan et al, ACL 2020)

https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

Tag and Generate

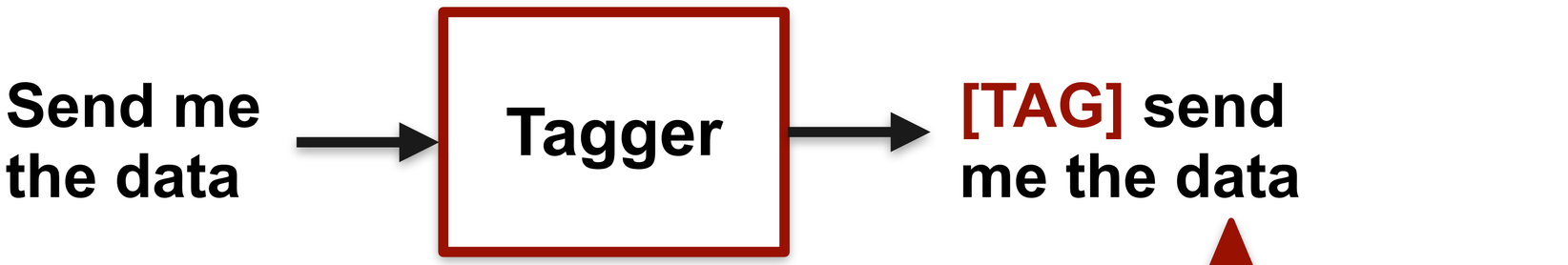
Tagger

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

(Madaan et al, ACL 2020)

https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

Tag and Generate



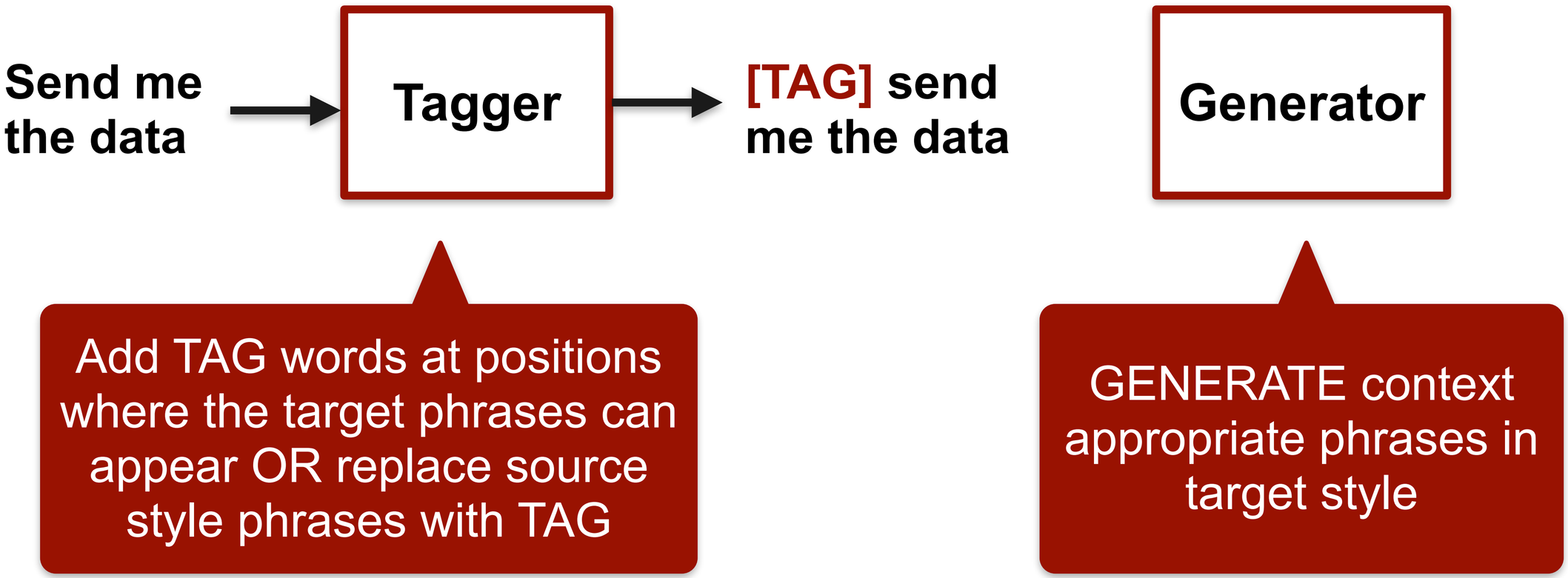
Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

This is representation z.

(Madaan et al, ACL 2020)

https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

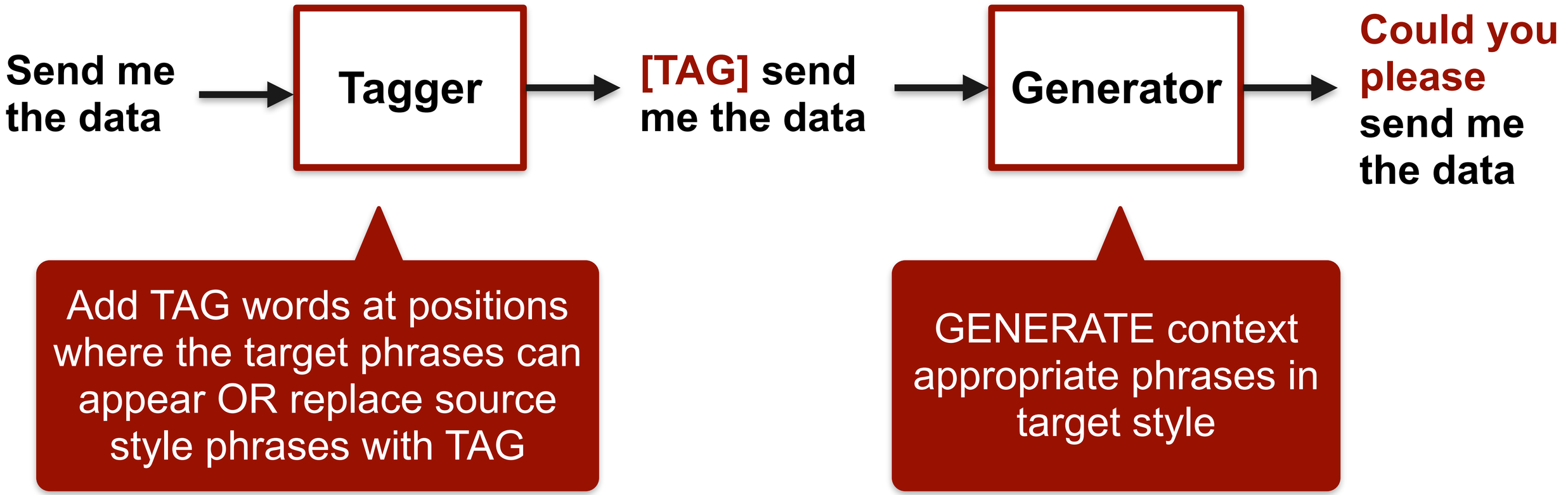
Tag and Generate



(Madaan et al, ACL 2020)

https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

Tag and Generate



(Madaan et al, ACL 2020)

https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

Tag and Generate

Tagger and Generator are transformers

Add TAG words at positions where the target phrases can appear OR replace source style phrases with TAG

GENERATE context appropriate phrases in target style

(Madaan et al, ACL 2020)

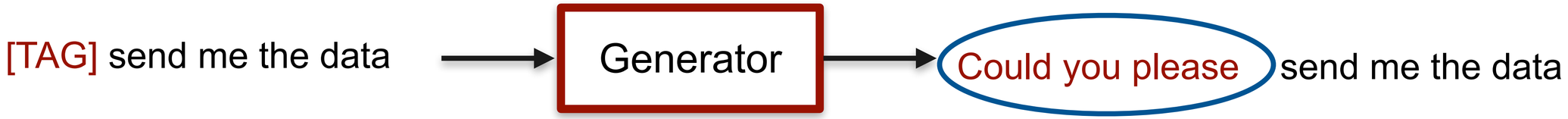
https://github.com/tag-and-generate/Politeness-Transfer-A-Tag-and-Generate-Approach

Goals achieved by Tag and Generate

- [ ] Transfer the Style
- [ ] Preserve the meaning
- [ ] Representation that reduces stylistic properties
- [ ] Interpretable Representation

Goals achieved by Tag and Generate

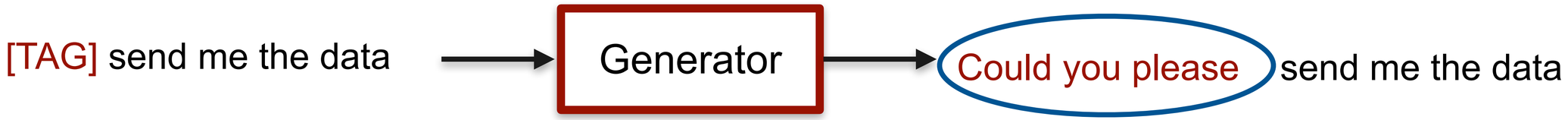
- [x] Transfer the Style



- [ ] Preserve the meaning
- [ ] Representation that reduces stylistic properties
- [ ] Interpretable Representation

Goals achieved by Tag and Generate

- [x] Transfer the Style



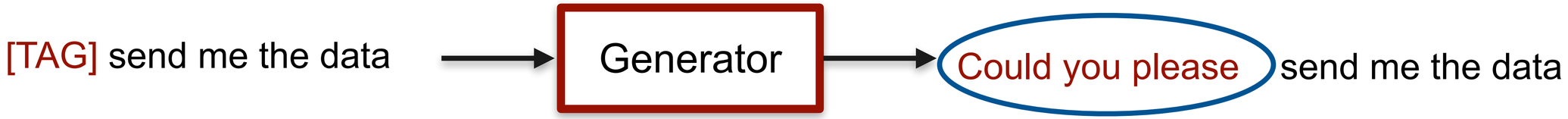
- [x] Preserve the meaning

- [ ] Representation that reduces stylistic properties

- [ ] Interpretable Representation

Goals achieved by Tag and Generate

- [x] Transfer the Style



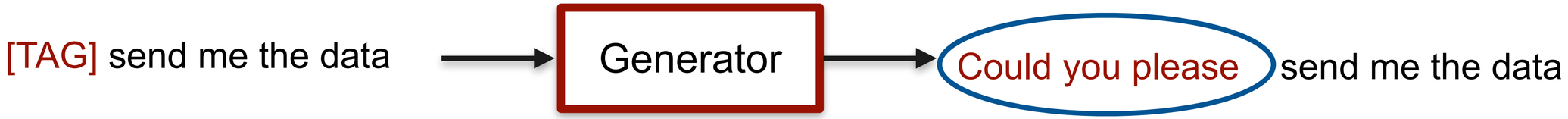
- [x] Preserve the meaning

- [x] Representation that reduces stylistic properties

- [ ] Interpretable Representation

Goals achieved by Tag and Generate

☑ Transfer the Style



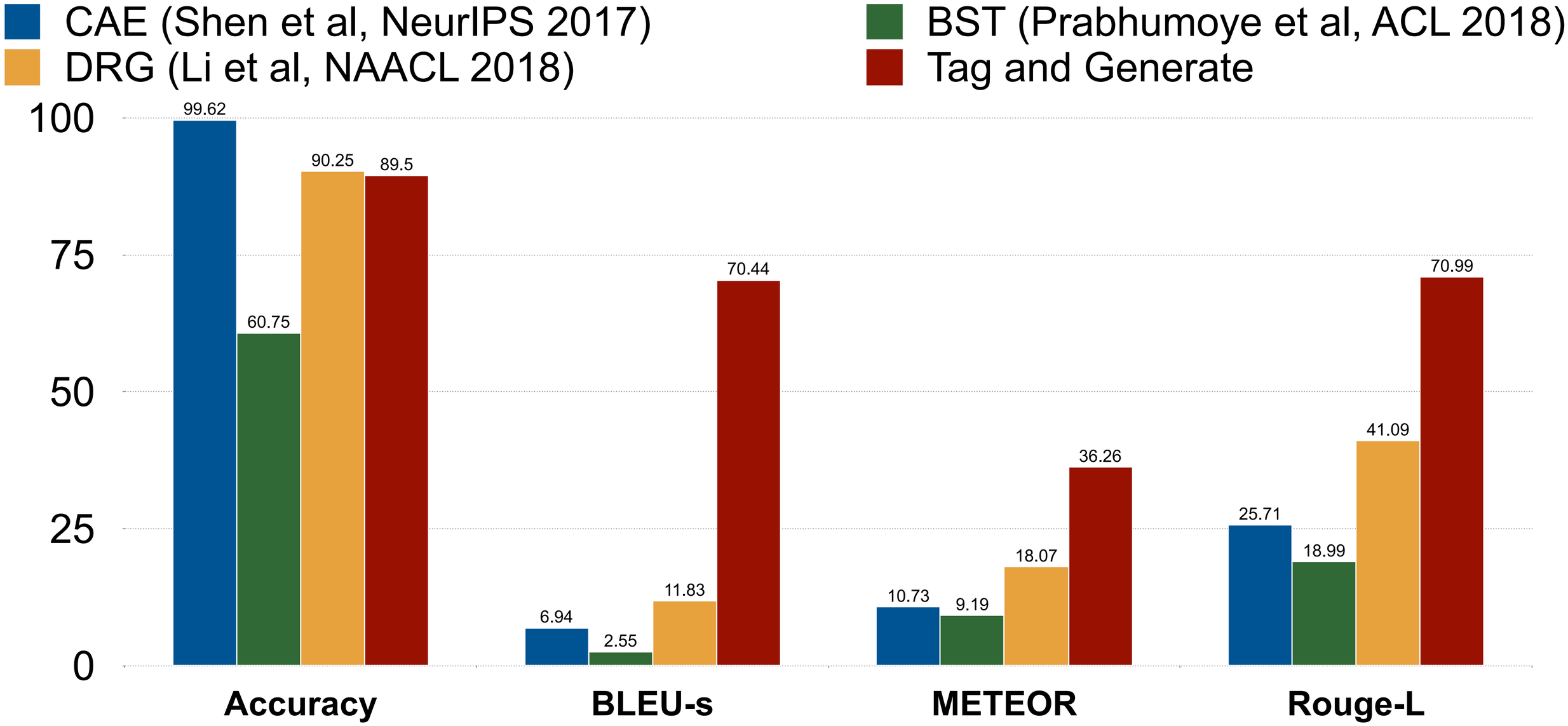
☑ Preserve the meaning

☑ Representation that reduces stylistic properties

☑ Interpretable Representation

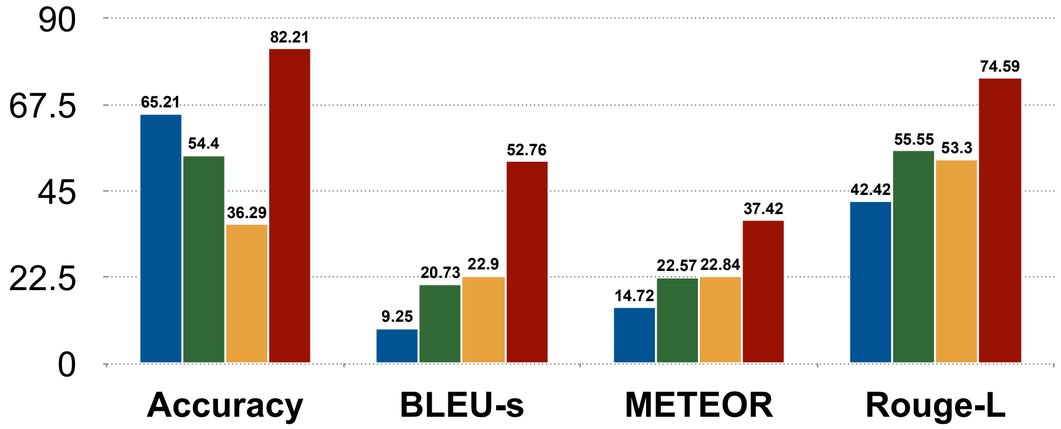
[TAG] send me the data

Politeness Results

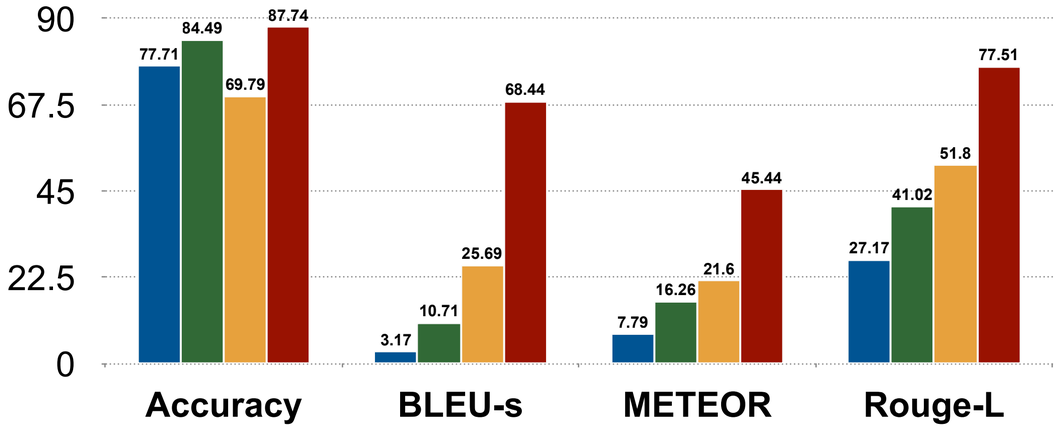


Results on four transfer tasks

Gender

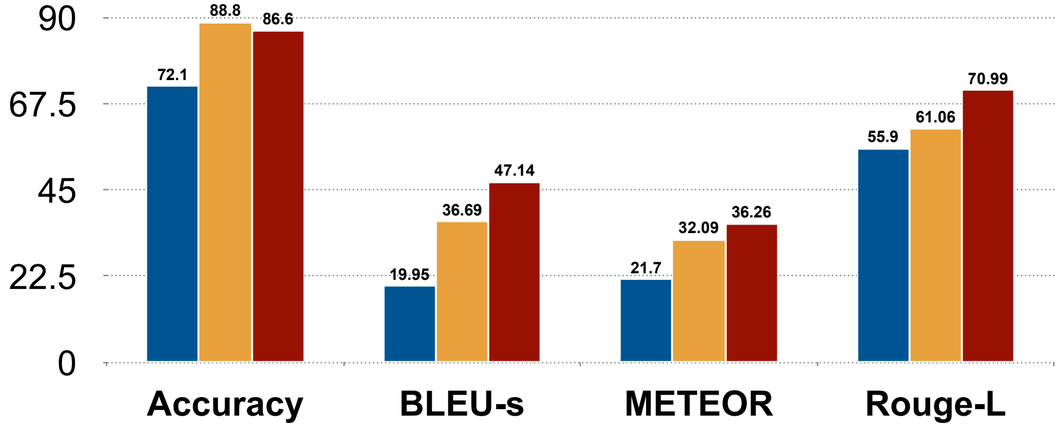


Political Slant

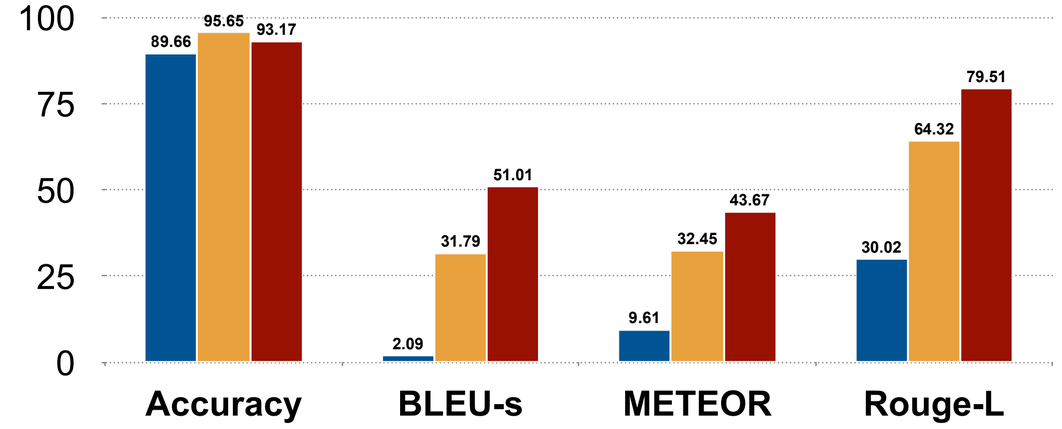


CAE (Shen et al, NeurIPS 2017) BST (Prabhumoye et al, ACL 2018) DRG (Li et al, NAACL 2018) Tag and Generate

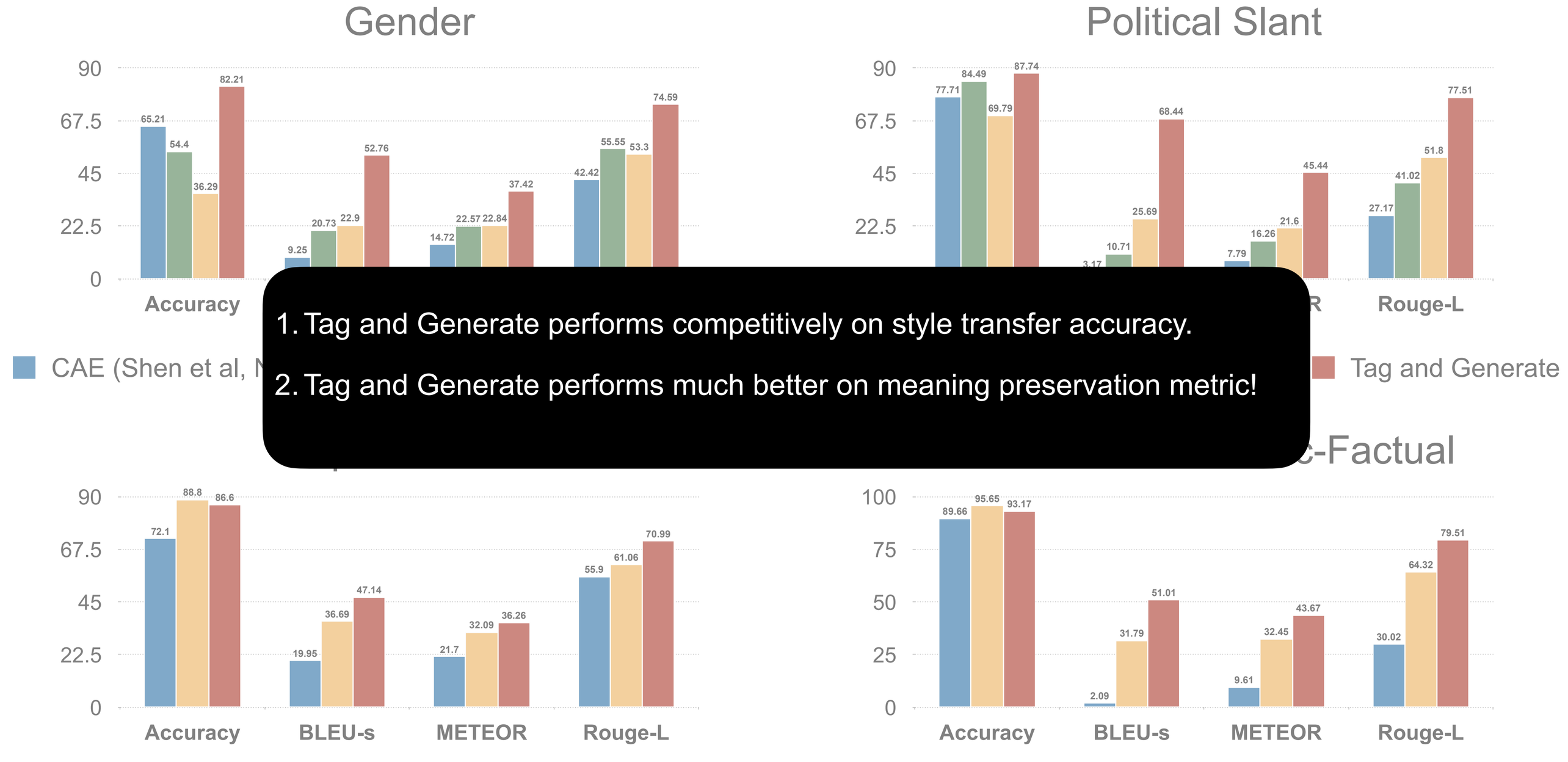
Yelp Sentiment



Humorous-Romantic-Factual



Results on four transfer tasks



Limitations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task

Limitations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task

Good drinks, and good company.

Positive to ***Negative***

Good drinks, and good company. ***But overall it sucked.***

Salsa is not hot or good.

Negative to ***Positive***

Salsa is not hot or good. ***But overall it was perfect.***

Limitations of automated metrics

- BLEU, METEOR, Rouge-L rely on n-gram overlap
- Classifiers can be fooled by polarizing keywords
- Naive Baseline on sentiment transfer task

Good drinks, and good company.

Positive to ***Negative***

Good drinks, and good company. ***But overall it sucked.***

Salsa is not hot or good.

Negative to ***Positive***

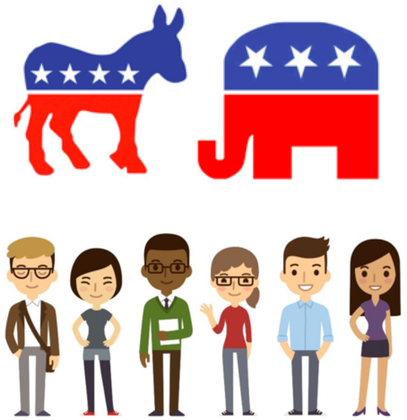
Salsa is not hot or good. ***But overall it was perfect.***

- 91.3% style transfer accuracy
- 61.44 BLEU score

Examples

Input	Output	Politeness Strategy
What happened to my personal station?	Could you please let me know what happened to my personal station?	**could** and **would** are forms which are polite
Yes, go ahead and remove it.	Yes, we can go ahead and remove it.	First person plural
Not yet - I'll try this weekend	Sorry, not yet - I'll try to make sure this weekend.	Apologizing

Overview



Style

ACL 2018
ACL 2020

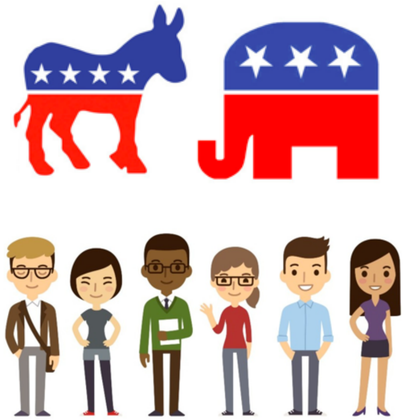
Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

ACL Findings
2022

Overview



Style

ACL 2018
ACL 2020

Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

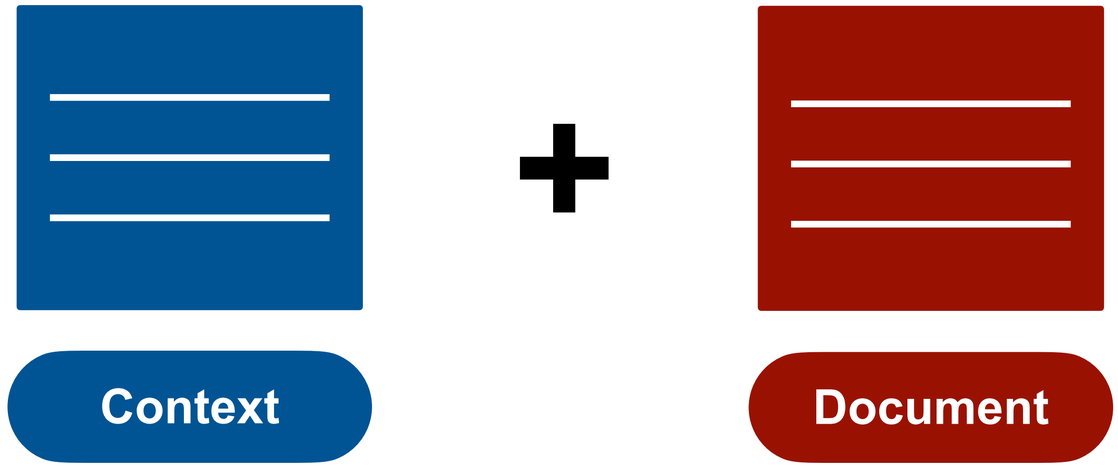
ACL Findings
2022

Document Grounded Generation

Document Grounded Generation

Context

Document Grounded Generation



Document Grounded Generation

Document Grounded Generation

- generate text given a context and a document
- generated text should *coherently fit* the context and contain new information from the document

Wikipedia Update Generation

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

Wikipedia Update Generation

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia

The **monkey selfie copyright dispute** is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog *Techdirt*, which have hosted the images following their publication in newspapers in July 2011 over Slater's objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a "selfie" picture might come about. The Wikimedia Foundation's 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

One of the monkey selfies at issue

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

Wikipedia Update Generation

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia

The **monkey selfie copyright dispute** is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog *Techdirt*, which have hosted the images following their publication in newspapers in July 2011 over Slater's objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a "selfie" picture might come about. The Wikimedia Foundation's 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

One of the monkey selfies at issue

Amateur Photographer

Subscribe now Digital

Latest Opinion Reviews How To Inspiration AP Awards The Video Mode

Home / Latest / Latest news

Ape-rture priority photographer plays down monkey reports

Chris Cheesman
July 5, 2011

0 shares

A photographer who says he witnessed monkeys taking pictures of themselves, tells Amateur Photographer (AP) that much of the media coverage has been exaggerated.

Wildlife photographer David Slater today played down newspaper reports that suggest a bunch of Indonesian monkeys grabbed his camera and began taking self-portraits.

And he revealed that the shots, featuring the endangered crested black macaque monkey, were taken three years ago.

Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

Latest reviews

Leica M10-R review

Andy Westlake takes a detailed look at Leica's latest high-resolution...

Canon EOS R6 review: hands-on first look

Look out Sony - Canon may just have got full-

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

Wikipedia Update Generation

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia

The **monkey selfie copyright dispute** is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog *Techdirt*, which have hosted the images following their publication in newspapers in July 2011 over Slater's objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a "selfie" picture might come about. The Wikimedia Foundation's 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

One of the monkey selfies at issue

Amateur Photographer

Subscribe now Digital

Latest Opinion Reviews How To Inspiration AP Awards The Video Mode

Home / Latest / Latest news

Ape-rture priority photographer plays down monkey reports

Chris Cheesman
July 5, 2011

0 shares

A photographer who says he witnessed monkeys taking pictures of themselves, tells Amateur Photographer (AP) that much of the media coverage has been exaggerated.

Wildlife photographer David Slater today played down newspaper reports that suggest a bunch of Indonesian monkeys grabbed his camera and began taking self-portraits.

And he revealed that the shots, featuring the endangered crested black macaque monkey, were taken three years ago.

Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

Latest reviews

Leica M10-R review

Andy Westlake takes a detailed look at Leica's latest high-resolution...

Canon EOS R6 review: hands-on first look

Look out Sony - Canon may just have got full-

The following day, *Amateur Photographer* reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[13]

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

Wikipedia Update Generation

WIKIPEDIA
The Free Encyclopedia

Not logged in Talk Contributions Create account Log in

Article Talk Read Edit View history Search Wikipedia

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia

The **monkey selfie copyright dispute** is a series of disputes about the copyright status of selfies taken by Celebes crested macaques using equipment belonging to the British nature photographer David Slater. The disputes involve Wikimedia Commons and the blog *Techdirt*, which have hosted the images following their publication in newspapers in July 2011 over Slater's objections that he holds the copyright, and People for the Ethical Treatment of Animals (PETA), who have argued that the macaque should be assigned the copyright.

Slater has argued that he has a valid copyright claim, as he engineered the situation that resulted in the pictures by travelling to Indonesia, befriending a group of wild macaques, and setting up his camera equipment in such a way that a "selfie" picture might come about. The Wikimedia Foundation's 2014 refusal to remove the pictures from its Wikimedia Commons image library was based on the understanding that copyright is held by the creator, that a non-human creator (not being a legal person) cannot hold copyright, and that the images are thus in the public domain.

Main page
Contents
Current events
Random article
About Wikipedia
Contact us
Donate

Contribute
Help
Community portal
Recent changes
Upload file

Tools
What links here
Related changes

One of the monkey selfies at issue

Amateur Photographer

Subscribe now Digital

Latest Opinion Reviews How To Inspiration AP Awards The Video Mode

Home / Latest / Latest news

Ape-rture priority photographer plays down monkey reports

Chris Cheesman
July 5, 2011

0 shares

A photographer who says he witnessed monkeys taking pictures of themselves, tells Amateur Photographer (AP) that much of the media coverage has been exaggerated.

Wildlife photographer David Slater today played down newspaper reports that suggest a bunch of Indonesian monkeys grabbed his camera and began taking self-portraits.

And he revealed that the shots, featuring the endangered crested black macaque monkey, were taken three years ago.

Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

Latest reviews

Leica M10-R review

Andy Westlake takes a detailed look at Leica's latest high-resolution...

Canon EOS R6 review: hands-on first look

Look out Sony - Canon may just have got full-

The following day, *Amateur Photographer* reported that Slater gave them further explanation as to how the photographs were created, downplaying the way newspaper articles had described them; Slater said reports that a monkey ran off with his camera and "began taking self-portraits" were incorrect and that the portrait was shot when his camera had been mounted on a tripod, with the primates playing around with a remote cable release as he fended off other monkeys.[13]

Generate text that is:

1. ***coherent*** to the Wikipedia context
2. contains ***new events*** from the article

Dataset Size of 636k!

(Prabhumoye et al, NAACL 2019)

https://github.com/shrimai/Towards-Content-Transfer-through-Grounded-Text-Generation

Document Grounded Dialogue

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG

Document Grounded Dialogue

Chat History

Alice: Yes this is a Christmas classic …

Bob: Very well said! Some of the critics were …

Alice: Yes I found the first half very entertaining.

Bob: I feel you! So hard to believe it was released in 1990!

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG

Document Grounded Dialogue

Chat History

Alice: Yes this is a Christmas classic …

Bob: Very well said! Some of the critics were …

Alice: Yes I found the first half very entertaining.

Bob: I feel you! So hard to believe it was released in 1990!

Article Talk

WIKIPEDIA
The Free Encyclopedia

Home Alone

From Wikipedia, the free encyclopedia

Home Alone is a 1990 American comedy film written and produced by John Hughes and directed by Chris Columbus. It stars Macaulay Culkin, Joe Pesci, Daniel Stern, John Heard, and Catherine O'Hara. Culkin stars as eight-year-old Kevin, who defends his home from burglars Harry and Marv (Pesci and Stern) after his family accidentally leaves him behind on their vacation.

Home Alone was filmed between February and May 1990 in locations throughout Illinois. It was originally set to be distributed by Warner Bros., but transferred to 20th Century Fox when Hughes exceeded Warner's $10 million budget.

Document

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG

Document Grounded Dialogue

Chat History

Alice: Yes this is a Christmas classic …

Bob: Very well said! Some of the critics were …

Alice: Yes I found the first half very entertaining.

Bob: I feel you! So hard to believe it was released in 1990!

Article Talk

WIKIPEDIA
The Free Encyclopedia

Home Alone

From Wikipedia, the free encyclopedia

Home Alone is a 1990 American comedy film written and produced by John Hughes and directed by Chris Columbus. It stars Macaulay Culkin, Joe Pesci, Daniel Stern, John Heard, and Catherine O'Hara. Culkin stars as eight-year-old Kevin, who defends his home from burglars Harry and Marv (Pesci and Stern) after his family accidentally leaves him behind on their vacation.

Home Alone was filmed between February and May 1990 in locations throughout Illinois. It was originally set to be distributed by Warner Bros., but transferred to 20th Century Fox when Hughes exceeded Warner's $10 million budget.

Document

Response

Alice: I like the quote "**sadistic festival of adult-bashing**". Between's **Kevin's parents forgetting him and the burglars** this movie doesn't make adult looks too bright.

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG

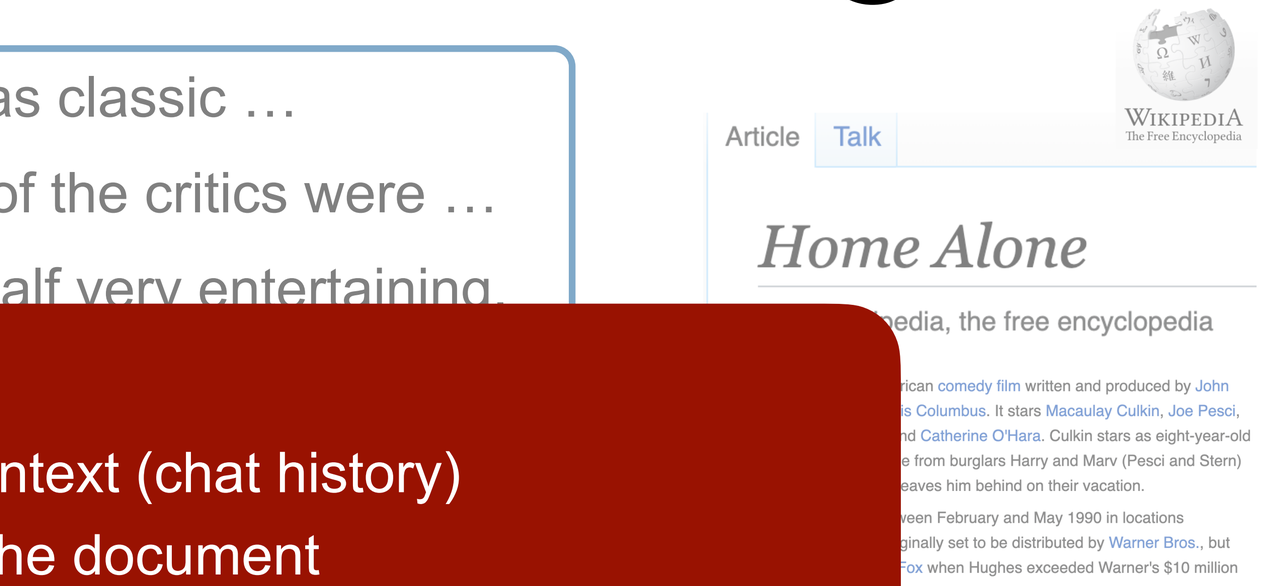
Document Grounded Dialogue

Alice: Yes this is a Christmas classic …

Bob: Very well said! Some of the critics were …

Alice: Yes I found the first half very entertaining.

Chat History



- Generate response that is:
  1. ***coherent*** to the dialogue context (chat history)
  2. contains ***information*** from the document
- CMU Document Grounded Conversations **(CMU_DoG)** Dataset
- Dataset size of ***130k utterances***!
- Wizard of Wikipedia (WoW) dataset has the same task.

Alice: I like the quote "**sadistic festival of adult-bashing**". Between's **Kevin's parents forgetting him and the burglars** this movie doesn't make adult looks too bright.

(Zhou, Prabhumoye & Black, EMNLP 2018)

https://github.com/festvox/datasets-CMU_DoG

Models

- Pre-trained encoder-decoder model
  - provide a stronger baseline for the task

Pre-trained Models

DOC

…

Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.

Pre-trained Models

DOC

…

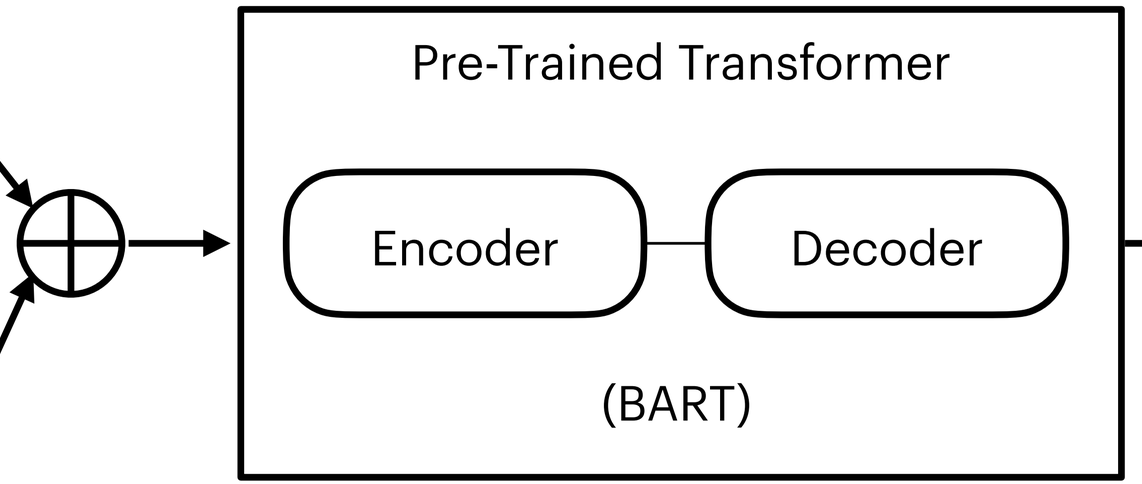
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



yes, that kind of union and solidarity is always needed in order to achieve anything.

Models

- Pre-trained encoder-decoder model
  - provide a stronger baseline for the task
- Context Driven Representation (CoDR)
  - build a representation of the content in document that is ***not present*** in the context
- Document Headed Attention (DoHA)
  - specific ***focus on document*** information

(Prabhumoye et al, NAACL 2021)

https://github.com/shrimai/Focused-Attention-Improves-Document-Grounded-Generation

Context Driven Representation (CoDR)

DOC

…

Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.

Context Driven Representation (CoDR)

DOC

…

Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.

Pre-Trained Transformer Encoder (BART Encoder)

$\mathbf{h}_d$

Context Driven Representation (CoDR)

DOC

…

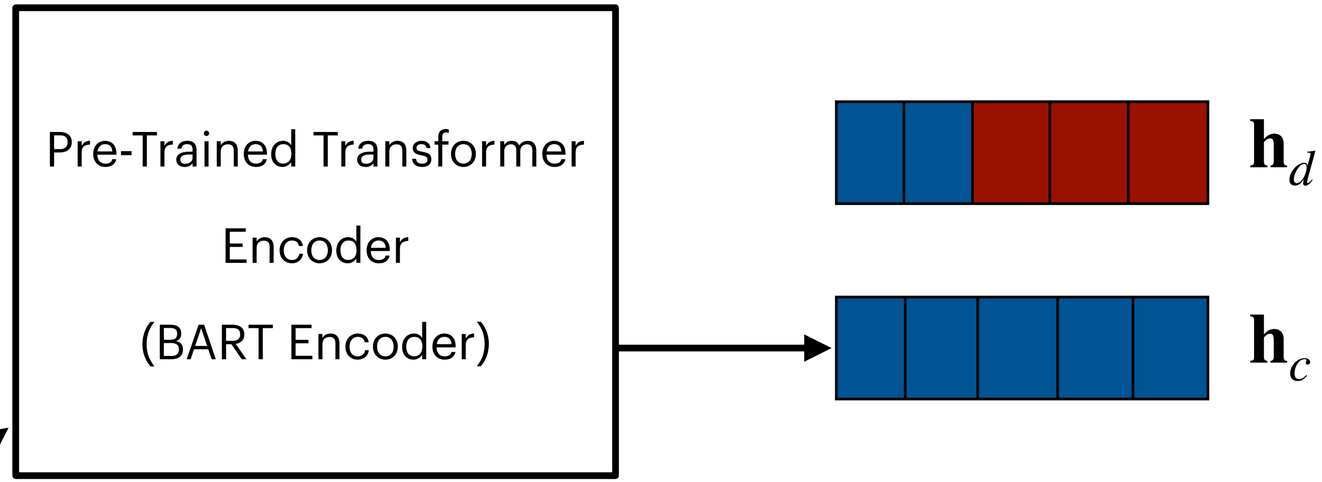
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



Context Driven Representation (CoDR)

DOC

…

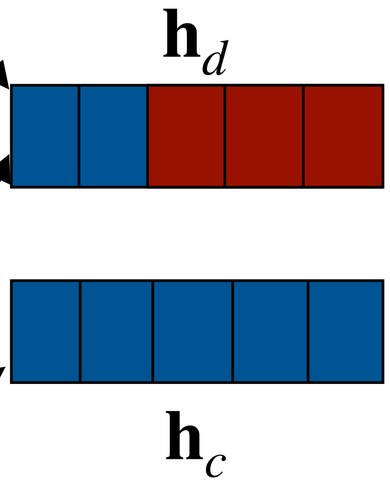
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



Context Driven Representation (CoDR)

DOC

…

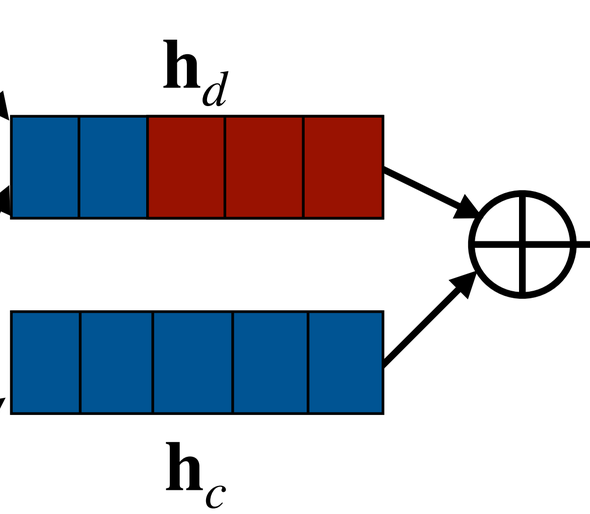
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



Pre-Trained Tranformer Decoder (BART Decoder)

Yes, but ***the newspapers won!***

Document Headed Attention (DoHA)

DOC

…

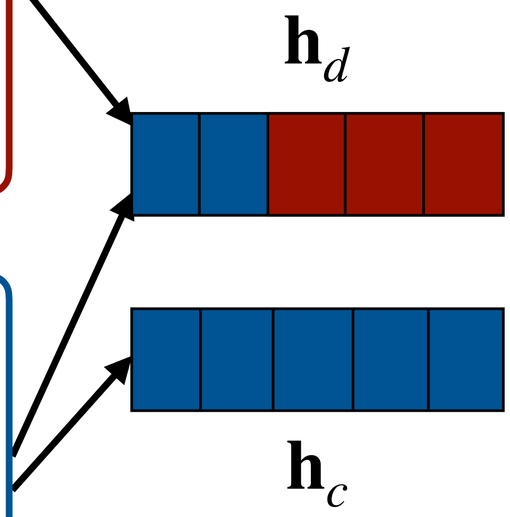
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

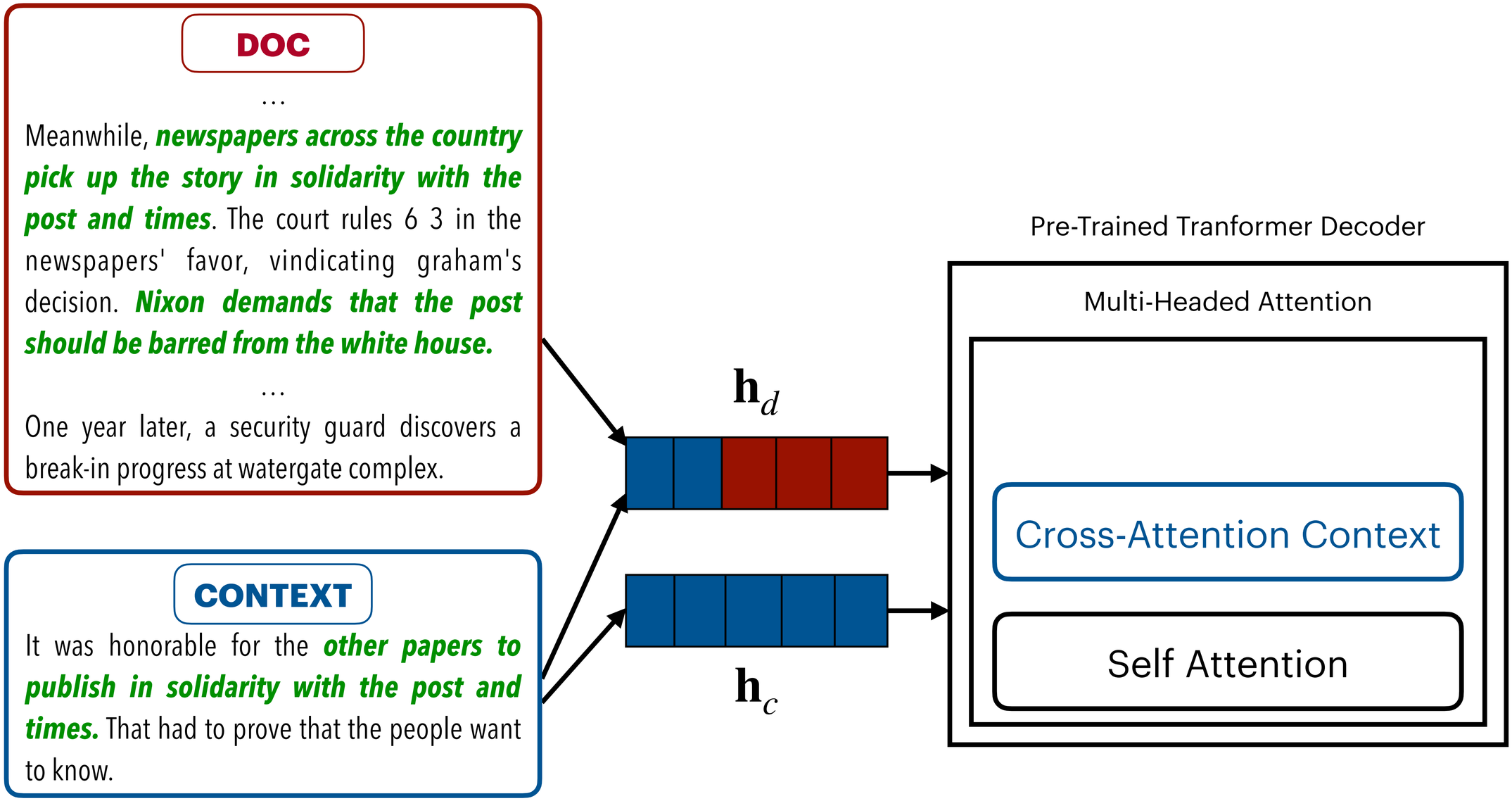
One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

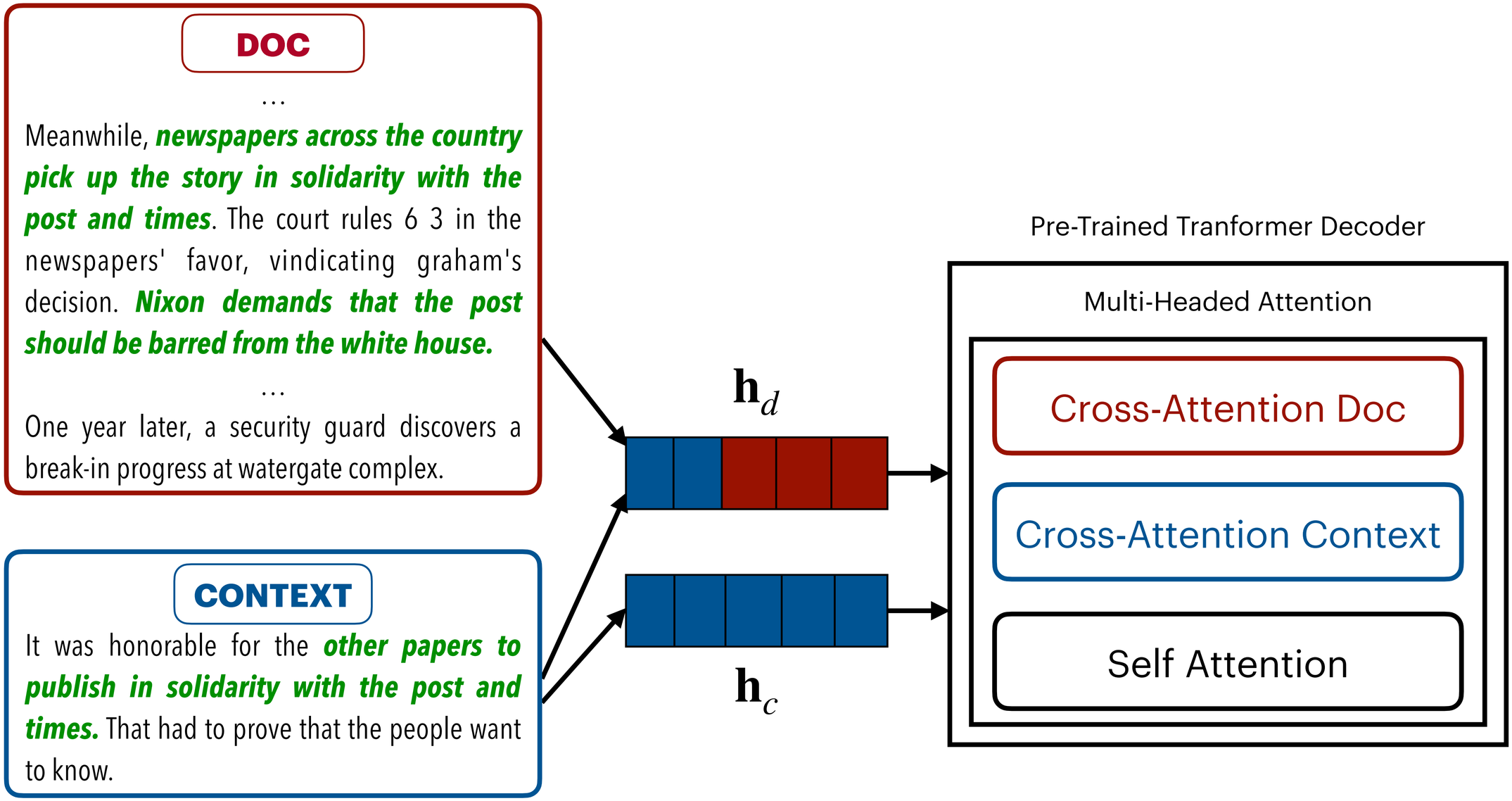
It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



Document Headed Attention (DoHA)



Document Headed Attention (DoHA)



Document Headed Attention (DoHA)

DOC

…

Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

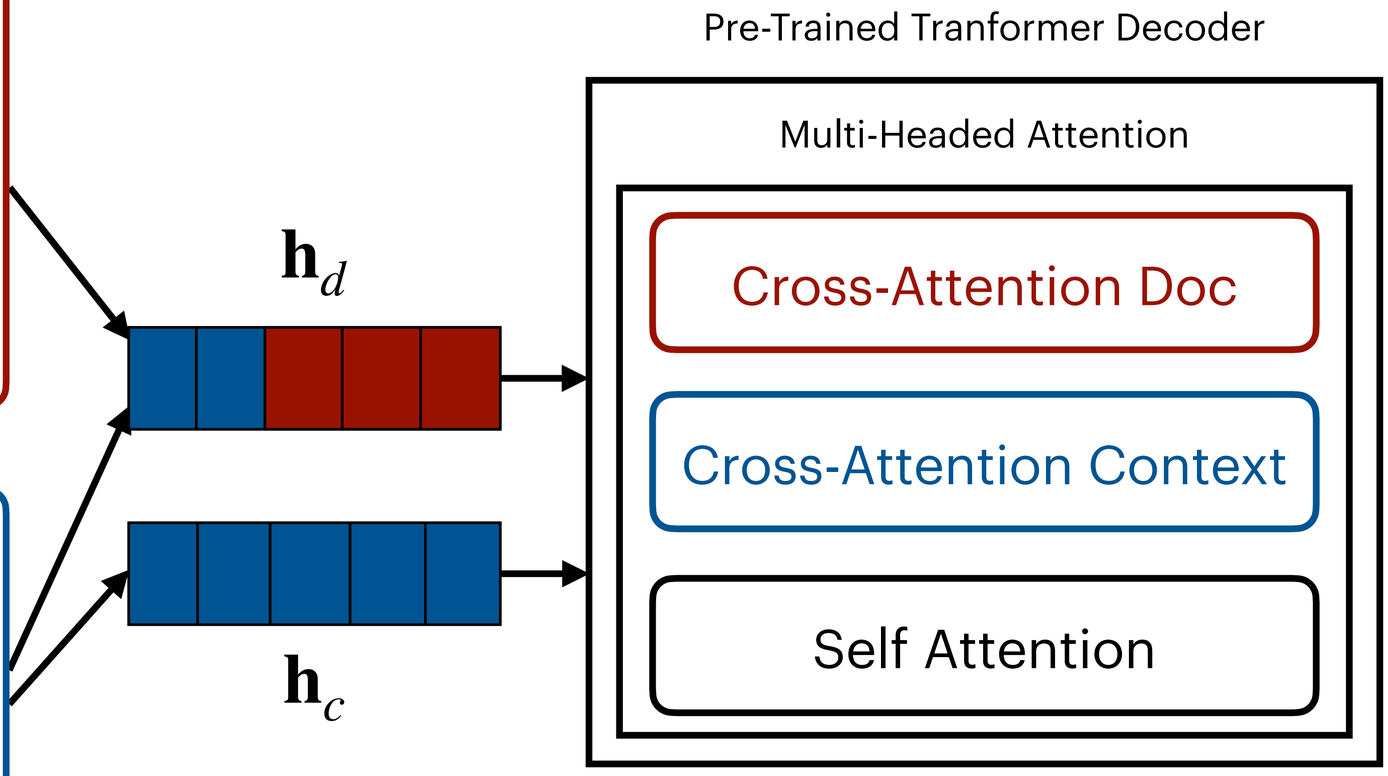
One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

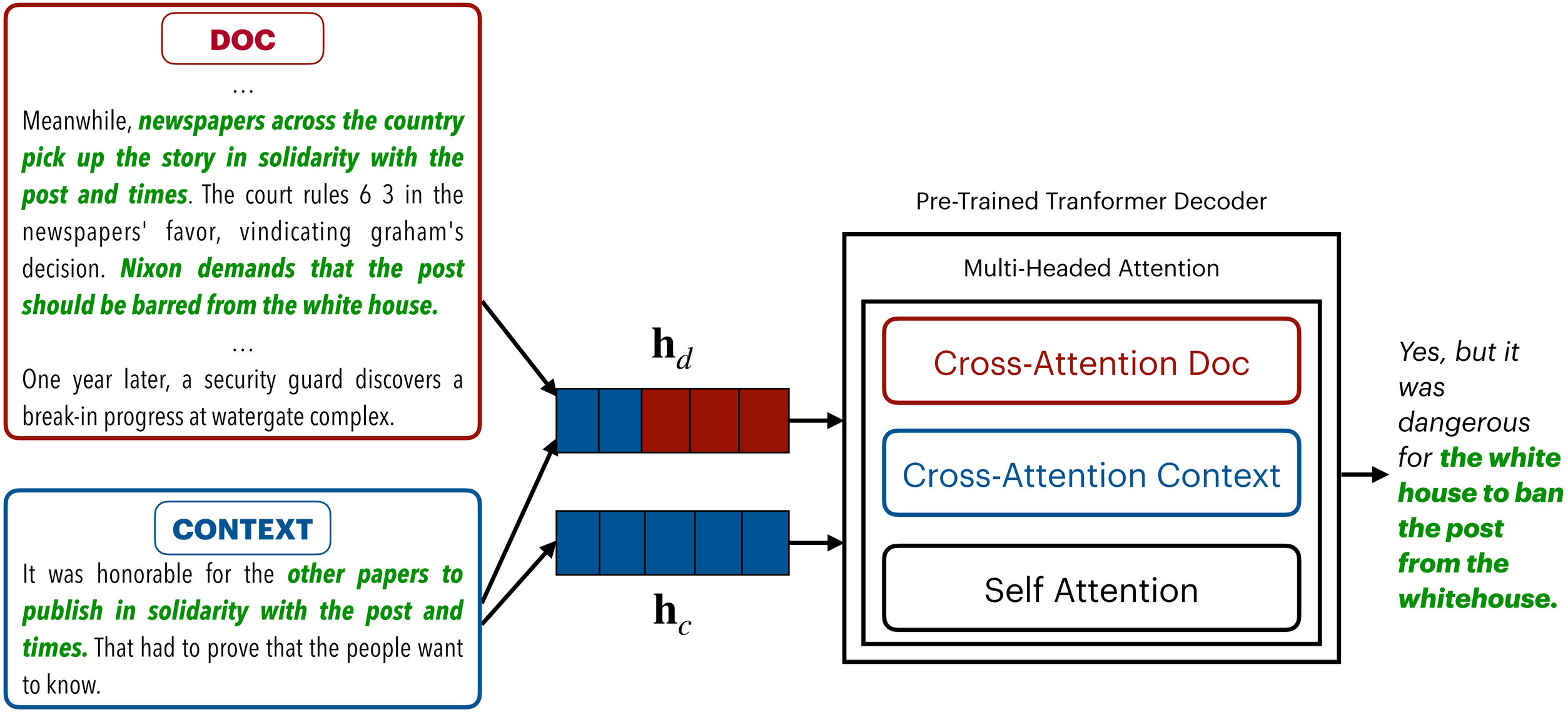
It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.

$$\texttt{CrossAttention_Doc}(Q, K, V) = [H_1, \ldots, H_m] \quad \mathbf{W}^{do}$$

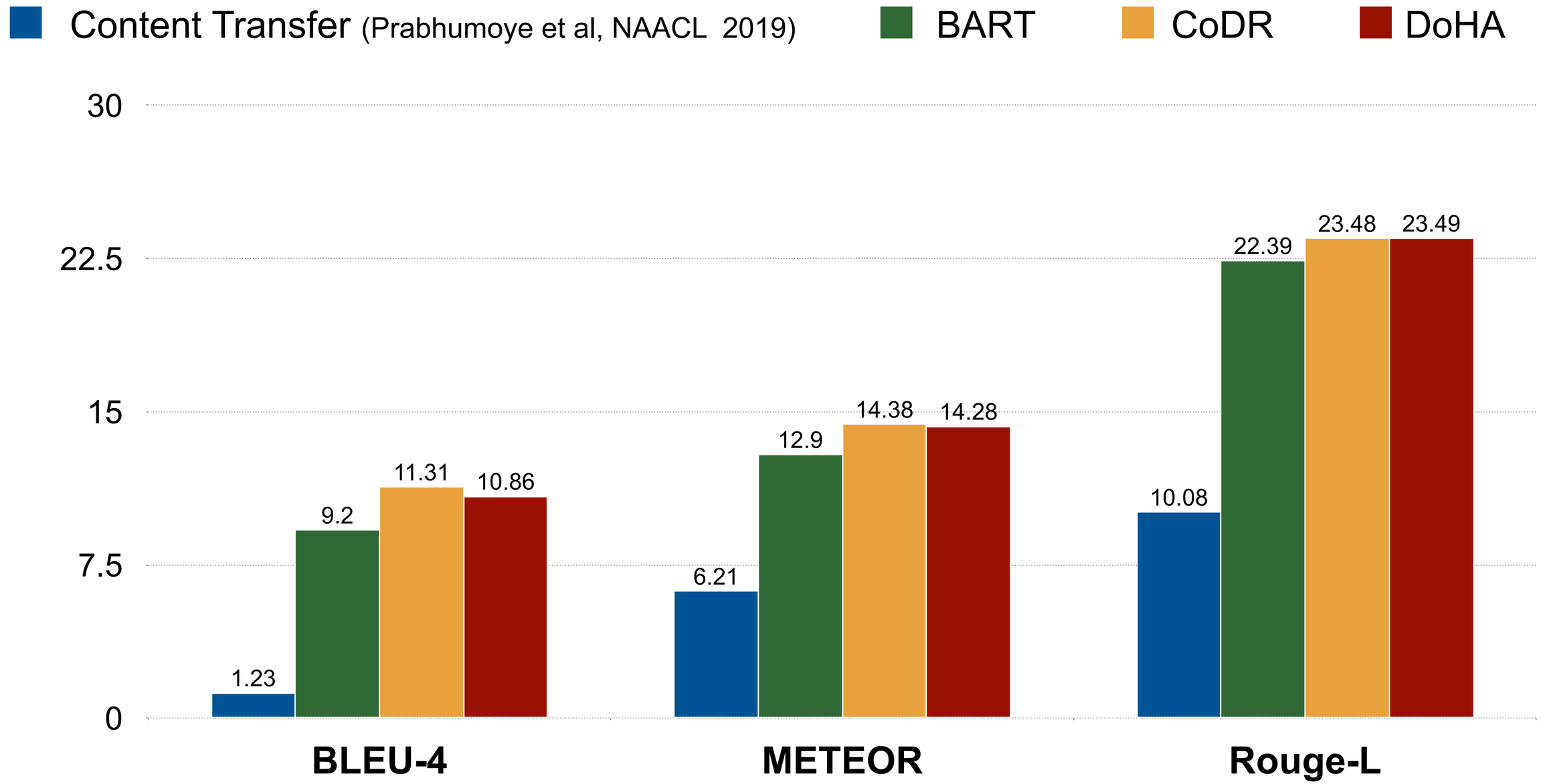
$$\mathbf{H}_i = \texttt{Attention}(Q\mathbf{W}_i^{dQ}, K\mathbf{W}_i^{dK}, V\mathbf{W}_i^{dV})$$



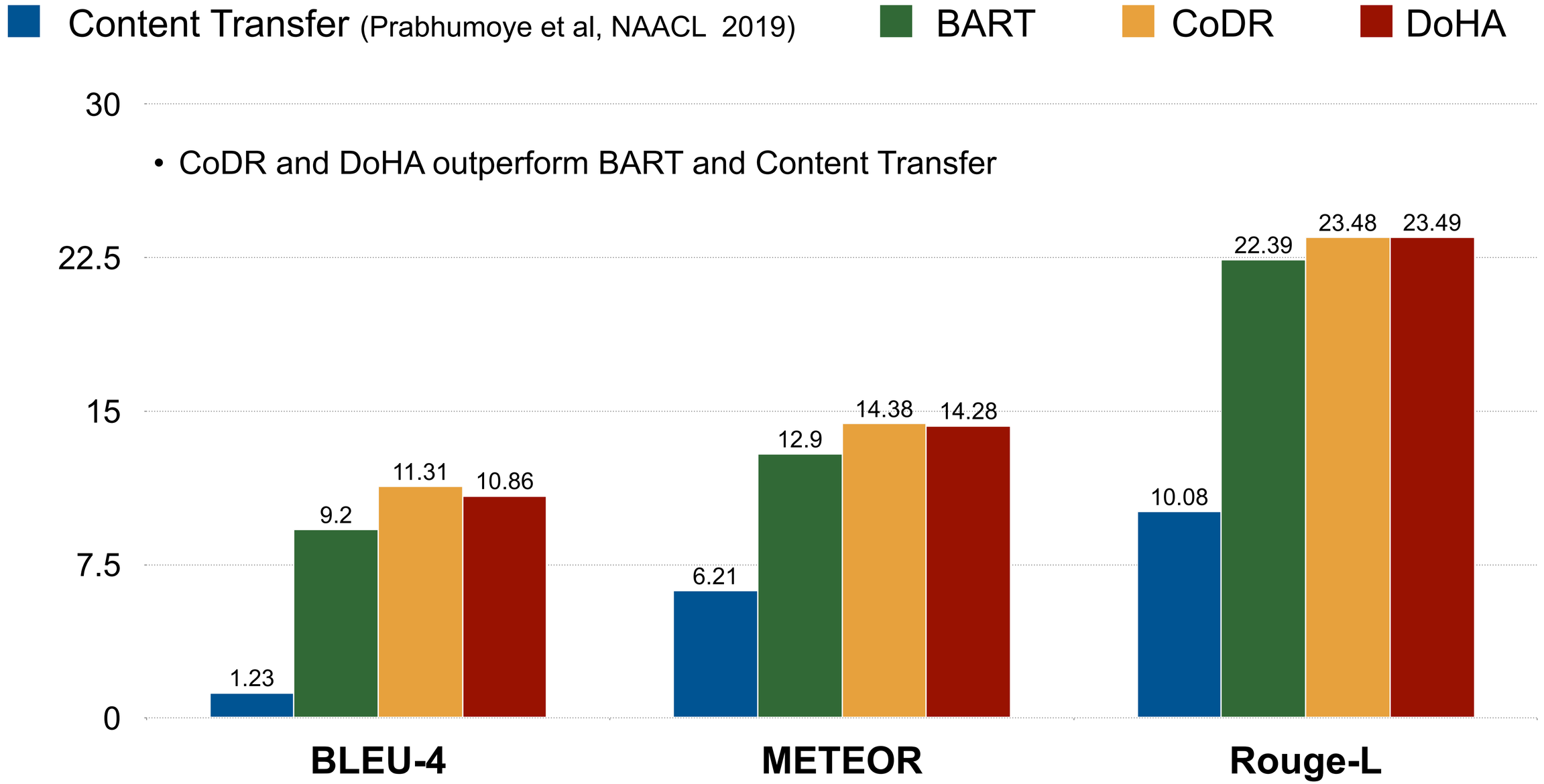
Document Headed Attention (DoHA)



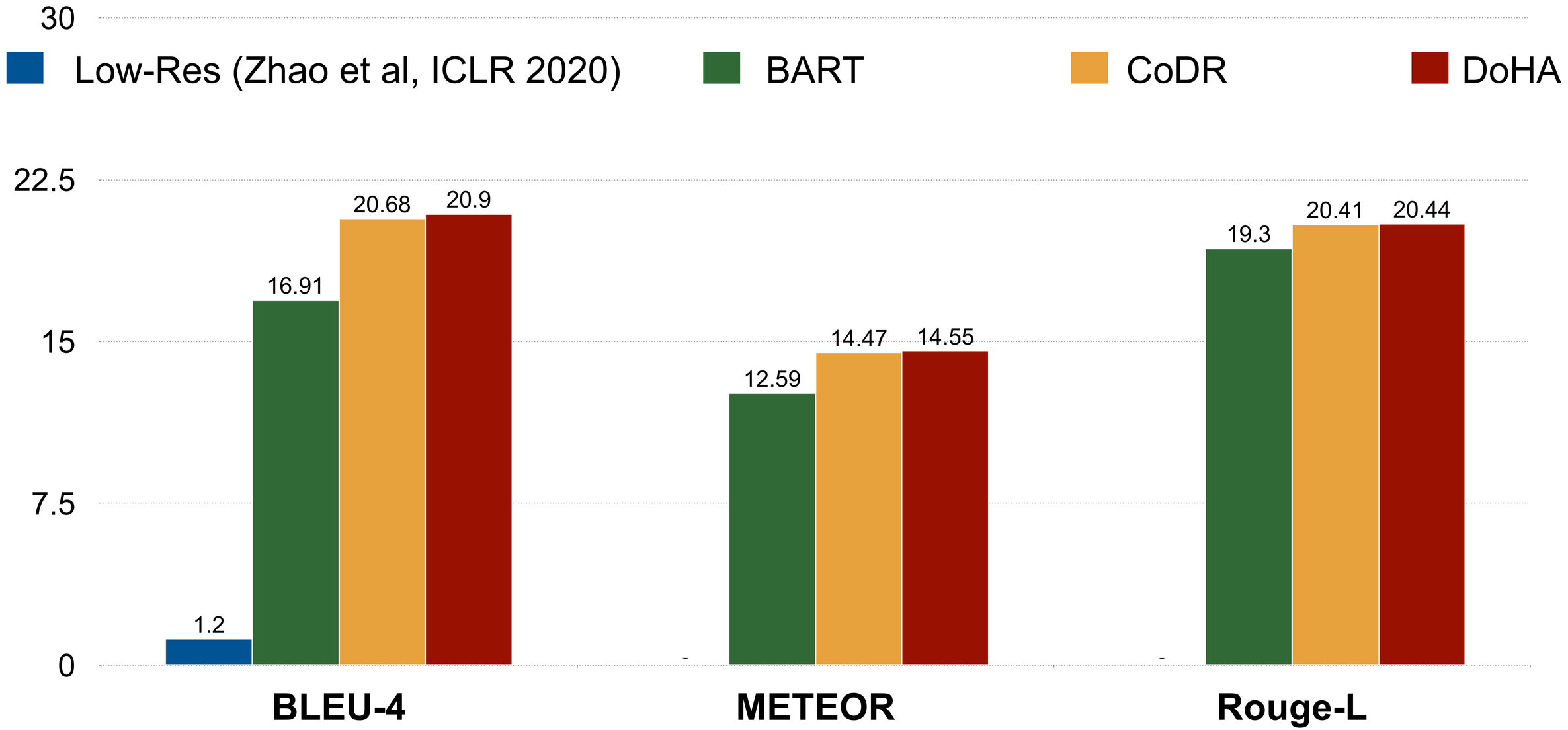
Wikipedia Update Generation



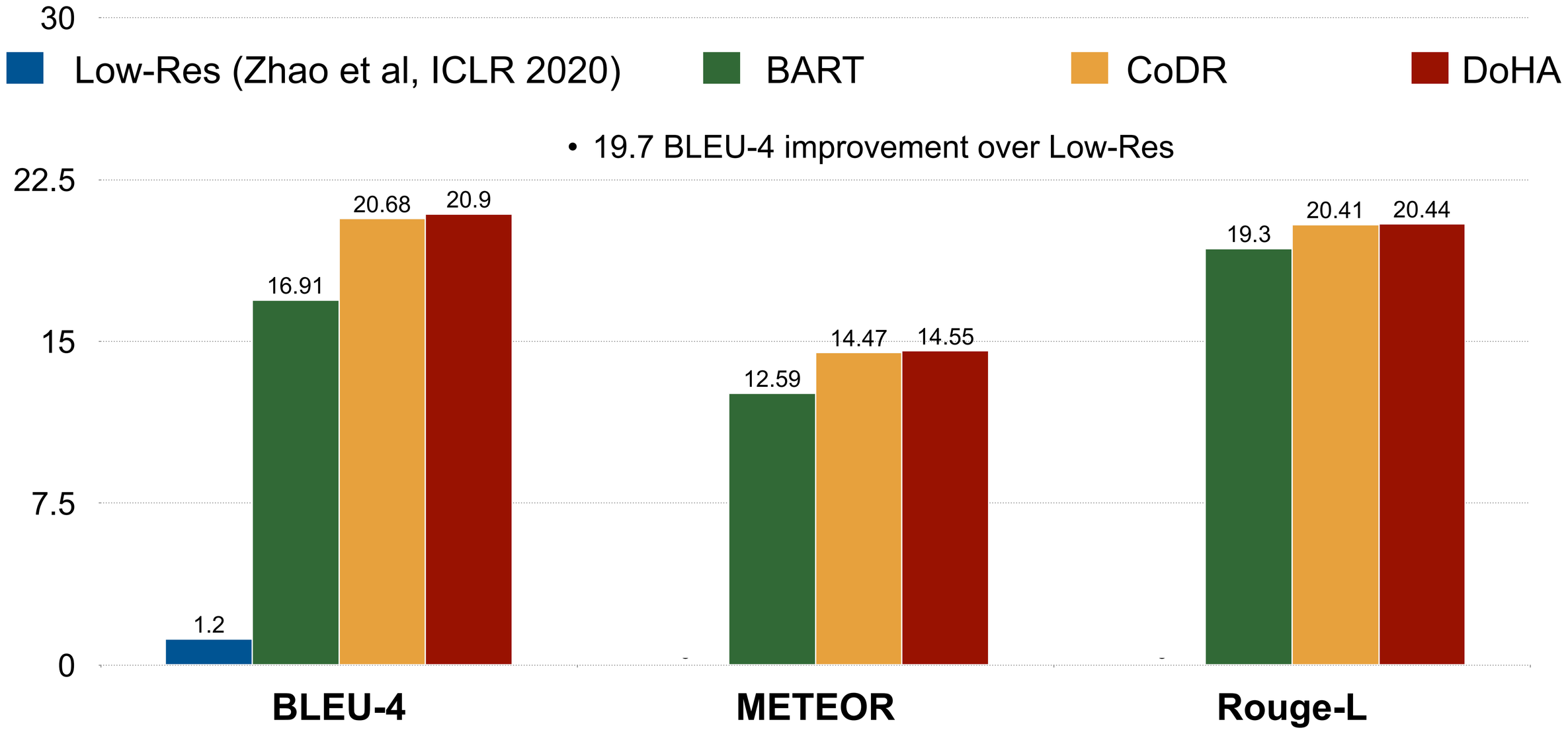
Wikipedia Update Generation



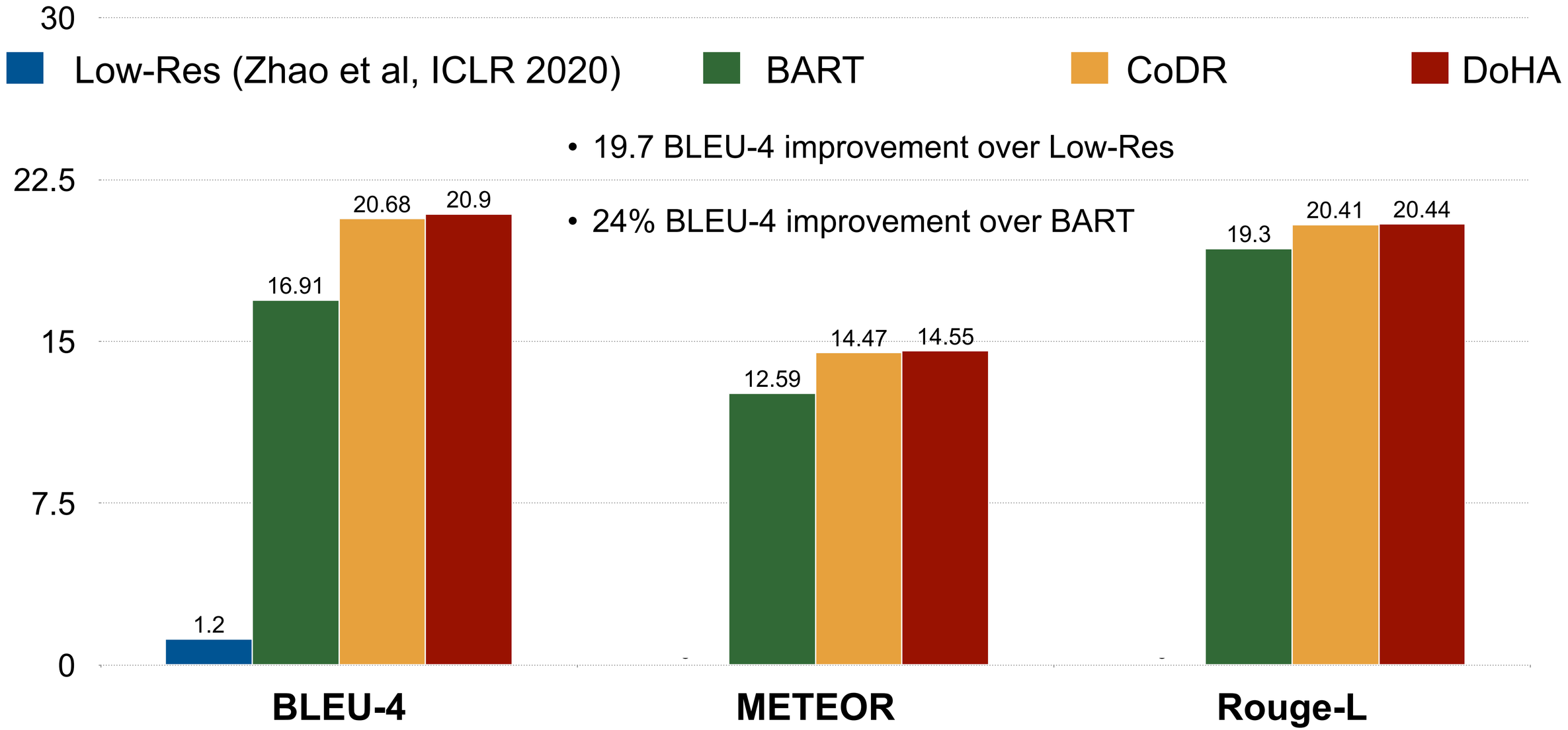
CMU_DoG


30
22.5
15
7.5
0
Low-Res (Zhao et al, ICLR 2020)
BART
CoDR
DoHA
1.2
16.91
20.68
20.9
12.59
14.47
14.55
19.3
20.41
20.44
BLEU-4
METEOR
Rouge-L

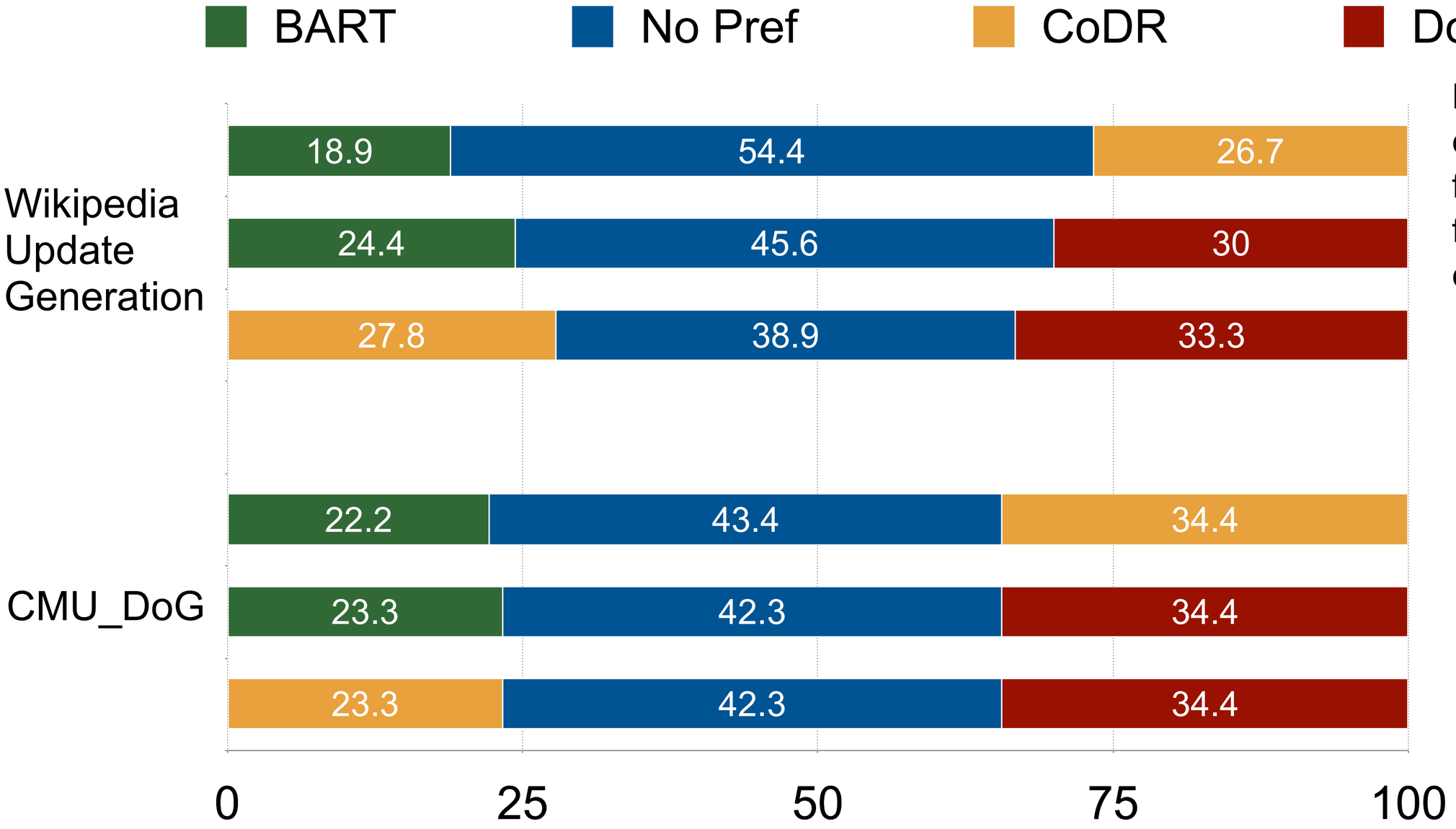
CMU_DoG


30
22.5
15
7.5
0
Low-Res (Zhao et al, ICLR 2020)
BART
CoDR
DoHA
• 19.7 BLEU-4 improvement over Low-Res
1.2
16.91
20.68
20.9
12.59
14.47
14.55
19.3
20.41
20.44
BLEU-4
METEOR
Rouge-L

CMU_DoG

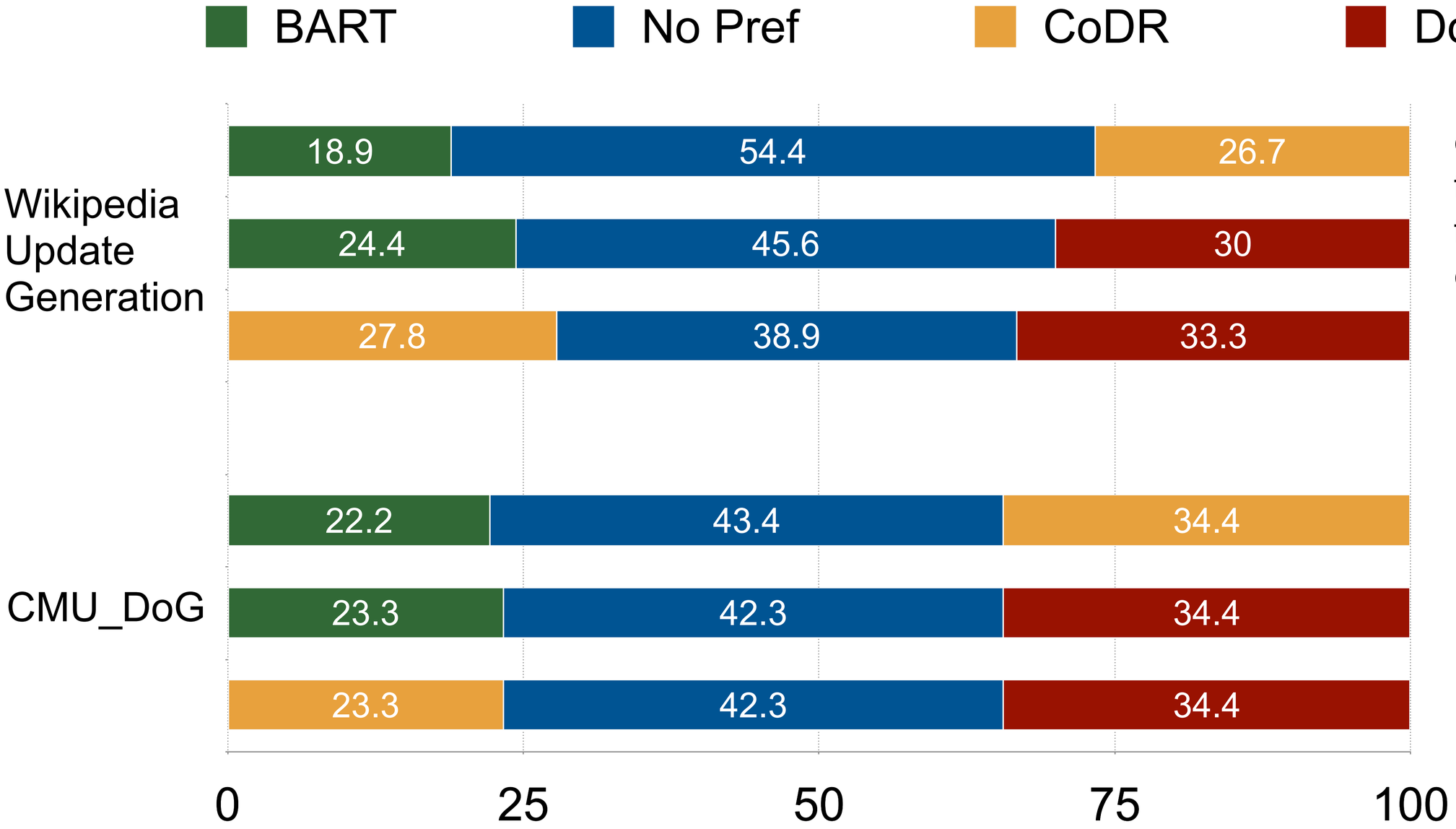

30
22.5
15
7.5
0
Low-Res (Zhao et al, ICLR 2020)
BART
CoDR
DoHA
• 19.7 BLEU-4 improvement over Low-Res
• 24% BLEU-4 improvement over BART
1.2
16.91
20.68
20.9
12.59
14.47
14.55
19.3
20.41
20.44
BLEU-4
METEOR
Rouge-L

Relevance Human Evaluation



Pick the option which contains information from the document and fits the context coherently

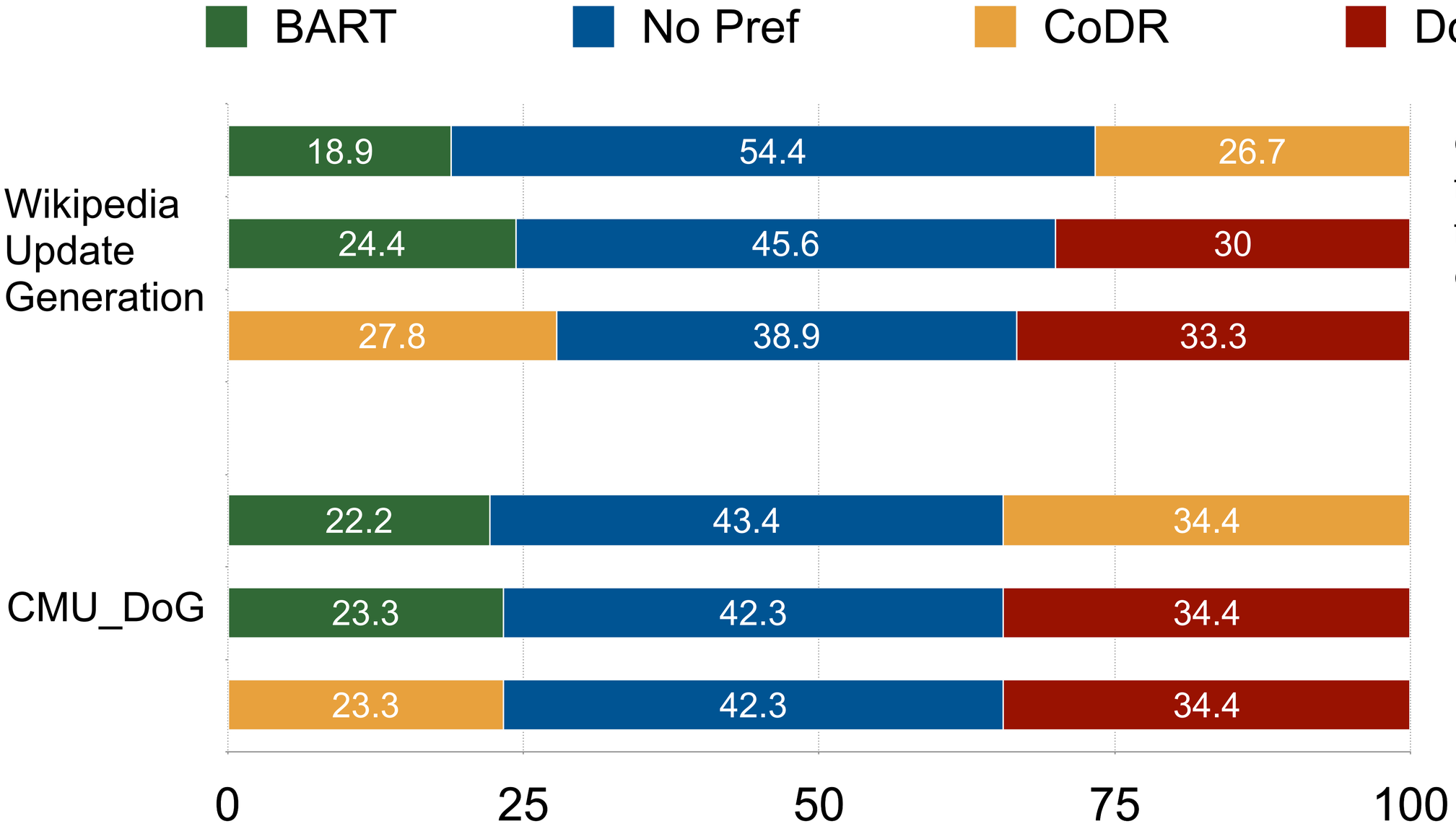
Relevance Human Evaluation



Pick the option which contains information from the document and fits the context coherently

- *No Preference is chosen most number of times.*

Relevance Human Evaluation



Pick the option which contains information from the document and fits the context coherently

- *No Preference is chosen most number of times.*
- *Both CoDR and DoHA outperform BART*

Wikipedia Update Generation

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41

Generation is a paraphrase of the Reference
Still gets a Rouge-L score of 41!
Inadequate automated metrics

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41
Partial Hallucination: Generation is either missing or hallucinates some information	23	America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.	The World Wide Web was first introduced on January 17, 1995 on Prodigy.	17

The generation has captured some information of the reference but missed some.

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41
Partial Hallucination: Generation is either missing or hallucinates some information	23	America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.	The World Wide Web was first introduced on January 17, 1995 on Prodigy.	17
Incoherent Reference: The reference does not coherently follow the context	22	“The Naked Ape” by Desmond Morris, is published.	Zoologist Desmond Morris publishes “The Naked Ape”.	26

Wikipedia pages that are in the form of lists
1340s, Timeline of DC Comics (1950s)
50% times the generation is grounded in the document and close to reference

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41
Partial Hallucination: Generation is either missing or hallucinates some information	23	America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.	The World Wide Web was first introduced on January 17, 1995 on Prodigy.	17
Incoherent Reference: The reference does not coherently follow the context	22	"The Naked Ape" by Desmond Morris, is published.	Zoologist Desmond Morris publishes "The Naked Ape".	26
Incorrect: The generation is either not appropriate or is not grounded (completely hallucinates the information).	7	The year 2000 is sometimes abbreviated as "Y2K" (the "Y" stands for "year", and the "K" stands for "kilo-" which means "thousand")	The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.	9

About 85% times, the generation is either completely or partially grounded if the reference is grounded

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41
Partial Hallucination: Generation is either missing or hallucinates some information	23	America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.	The World Wide Web was first introduced on January 17, 1995 on Prodigy.	17
Incoherent Reference: The reference does not coherently follow the context	22	"The Naked Ape" by Desmond Morris, is published.	Zoologist Desmond Morris publishes "The Naked Ape".	26
Incorrect: The generation is either not appropriate or is not grounded (completely hallucinates the information).	7	The year 2000 is sometimes abbreviated as "Y2K" (the "Y" stands for "year", and the "K" stands for "kilo-" which means "thousand")	The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.	9
Reference is not grounded	5	This was achieved under dead calm conditions as an additional safety measure, whereas the Wrights flew in a 25 mph+ wind to achieve enough airspeed on their early attempts.	This was verified by a video crew present at the test flight.	14

Dataset Noise!

Wikipedia Update Generation

Error Class	%	Reference	Generation	R
Linguistic Variation: Generation is appropriate but a linguistic variation of the reference.	43	December 12 - The Smiths play Brixton Academy, their last ever gig before their dissolution.	December 12 - The Smiths perform their final show, at Brixton Academy in London.	41
Partial Hallucination: Generation is either missing or hallucinates some information	23	America Online and Prodigy (online service) offered access to the World Wide Web system for the first time this year, releasing browsers that made it easily accessible to the general public.	The World Wide Web was first introduced on January 17, 1995 on Prodigy.	17
Incoherent Reference: The reference does not coherently follow the context	22	"The Naked Ape" by Desmond Morris, is published.	Zoologist Desmond Morris publishes "The Naked Ape".	26
Incorrect: The generation is either not appropriate or is not grounded (completely hallucinates the information).	7	The year 2000 is sometimes abbreviated as "Y2K" (the "Y" stands for "year", and the "K" stands for "kilo-" which means "thousand")	The Y2K conspiracy theory claimed that a secret nuclear attack by the United States on 2 January 2000 was planned to begin World War 2.	9
Reference is not grounded	5	This was achieved under dead calm conditions as an additional safety measure, whereas the Wrights flew in a 25 mph+ wind to achieve enough airspeed on their early attempts.	This was verified by a video crew present at the test flight.	14

Observations on CMU_DoG

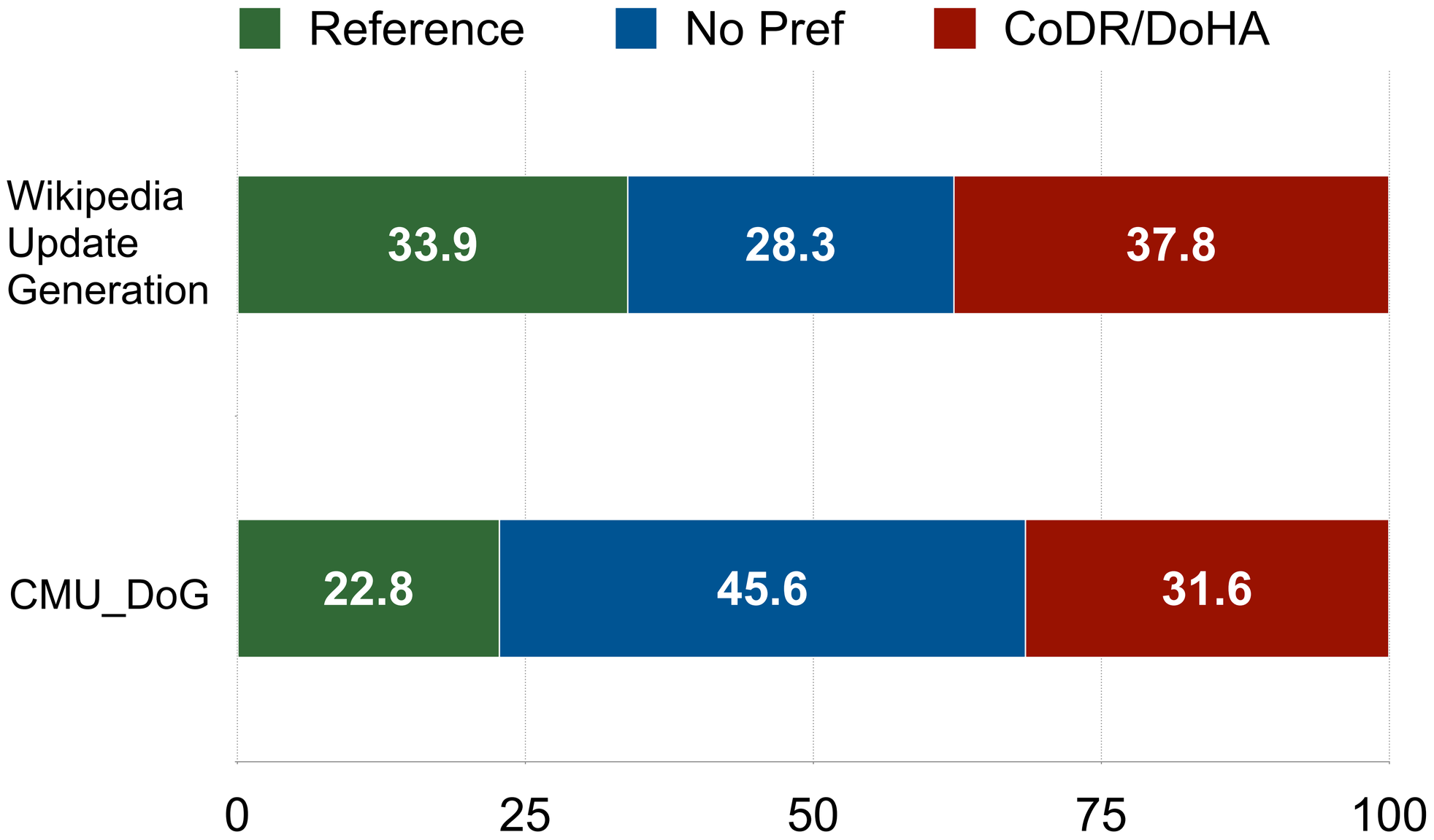
Error Class	%
Reference and generation are grounded	35
Reference is grounded but generation is not	3
Reference is not grounded but generation is	5
Reference and generation are not grounded	43

Observations on CMU_DoG

Error Class	%
Reference and generation are grounded	35
Reference is grounded but generation is not	3
Reference is not grounded but generation is	5
Reference and generation are not grounded	43

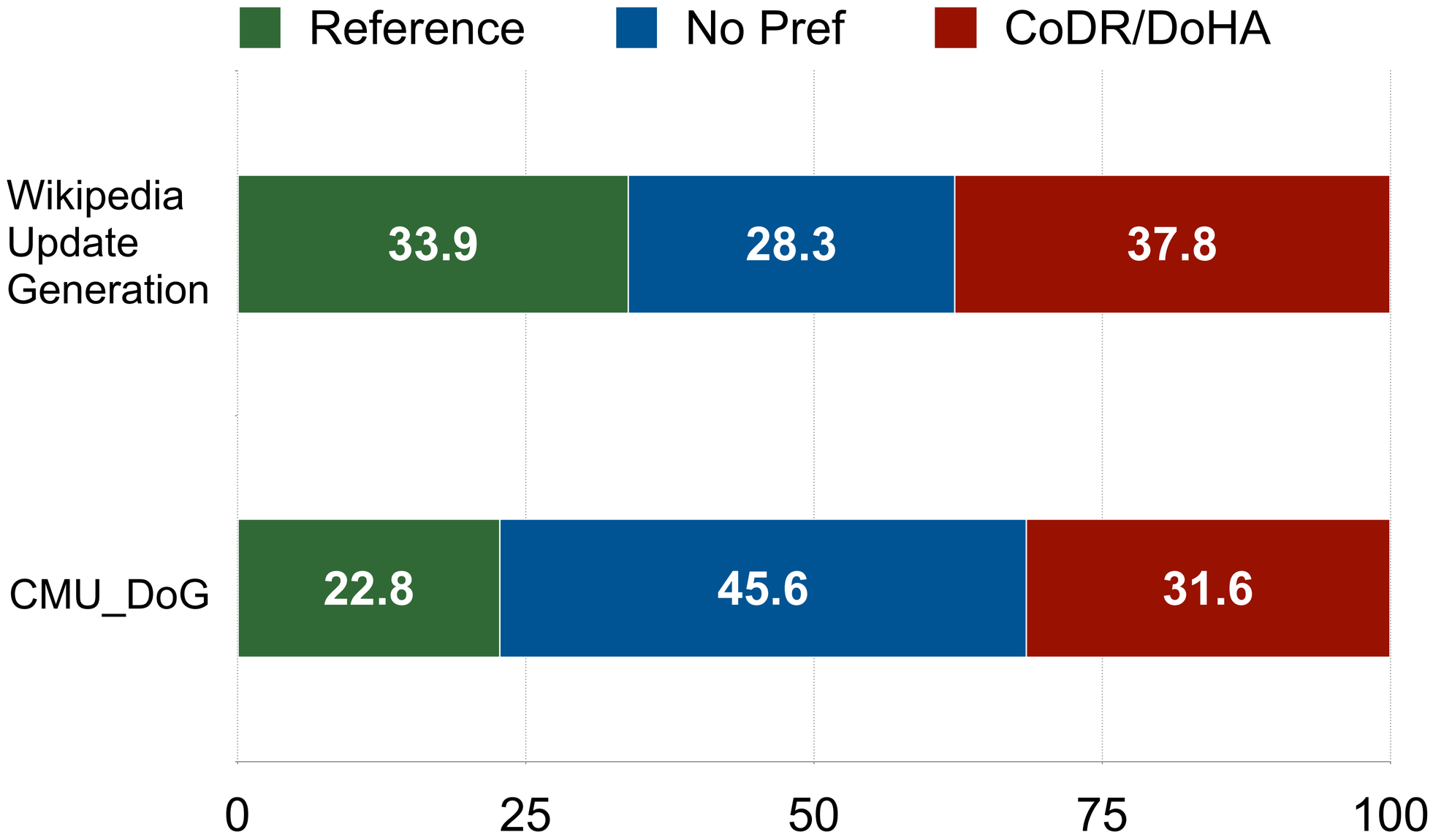
1. References are not grounded if they are follow up ***questions, opinions or experiences.***
2. ***Context dictates*** grounding
3. ***Conversational subtleties not captured*** by the automated metrics
4. ***90%*** times the generation is grounded if the reference is grounded and generation is not grounded if the reference is not grounded
5. ***86.7%*** generated responses are ***appropriate to the context!***

Reference vs CoDR/DoHA



Pick the option that is most appropriate to the given context.

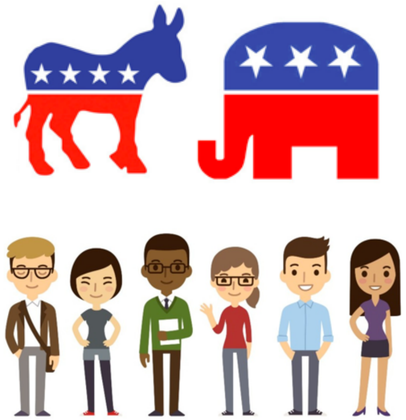
Reference vs CoDR/DoHA



Pick the option that is most appropriate to the given context.

1. Automated metrics are inadequate
2. Sole reference should not be considered as the only correct response to the context

Overview



Style

ACL 2018
ACL 2020

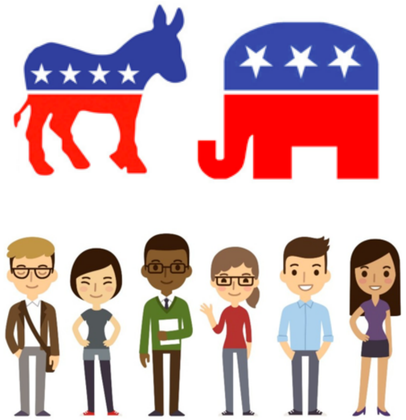
Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

ACL Findings
2022

Overview



Style

ACL 2018
ACL 2020

Content

EMNLP 2018
NAACL 2019
NAACL 2021

LMs

ACL Findings
2022

Controllable Text Generation

Using Pre-trained Language Models

- Language generation is now fluent and grammatical!
- How useful are LMs in steering the output in the desired direction?
- Harness the knowledge from LMs for QA tasks and classification tasks
  - Can it be used for controllable text generation tasks?

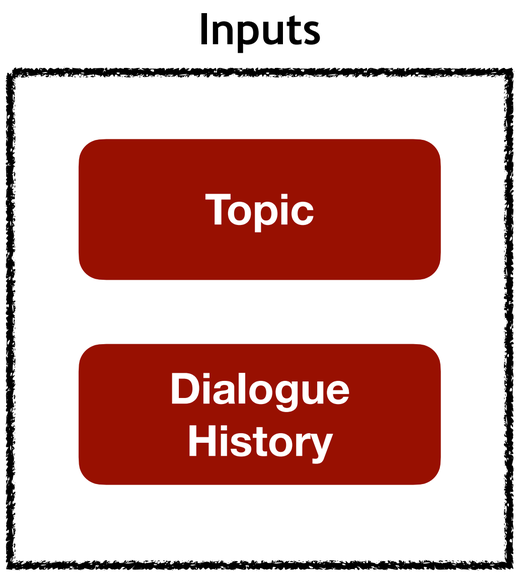
Challenges in building Large Language Models

- High compute requirements
  - Made efficient due to NVLink, NVSwitch and Mellanox 200Gbps Infiniband cards
- Model Parameters don’t fit in GPU memory
  - Efficient and scalable 3D parallel system capable of combining data, pipeline and tensor-slicing based parallelism
- Evaluations
  - Benchmark testing to understand bias and toxicity
  - NLP benchmark tests may not be enough

Multi-Stage Dialogue Prompting Framework (MSDP)

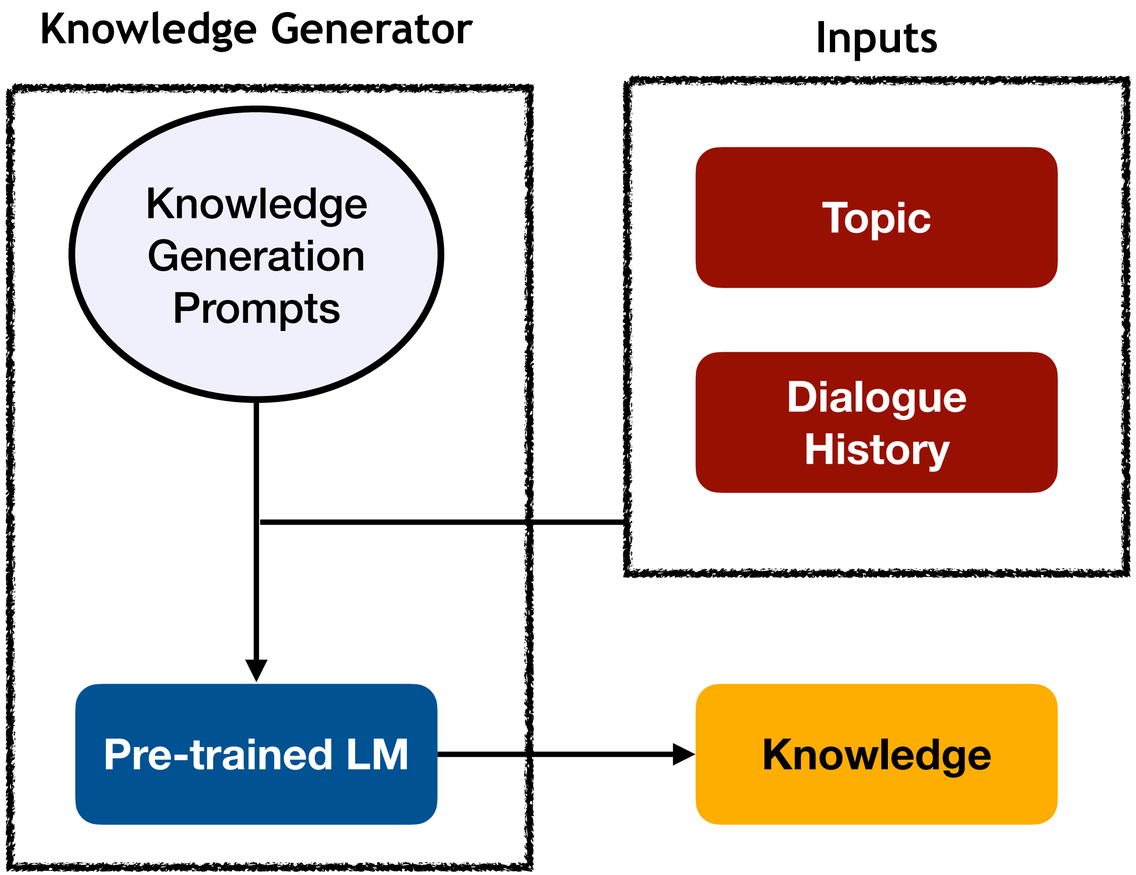
(Liu et al, ACL Findings 2022)

Multi-Stage Dialogue Prompting Framework (MSDP)



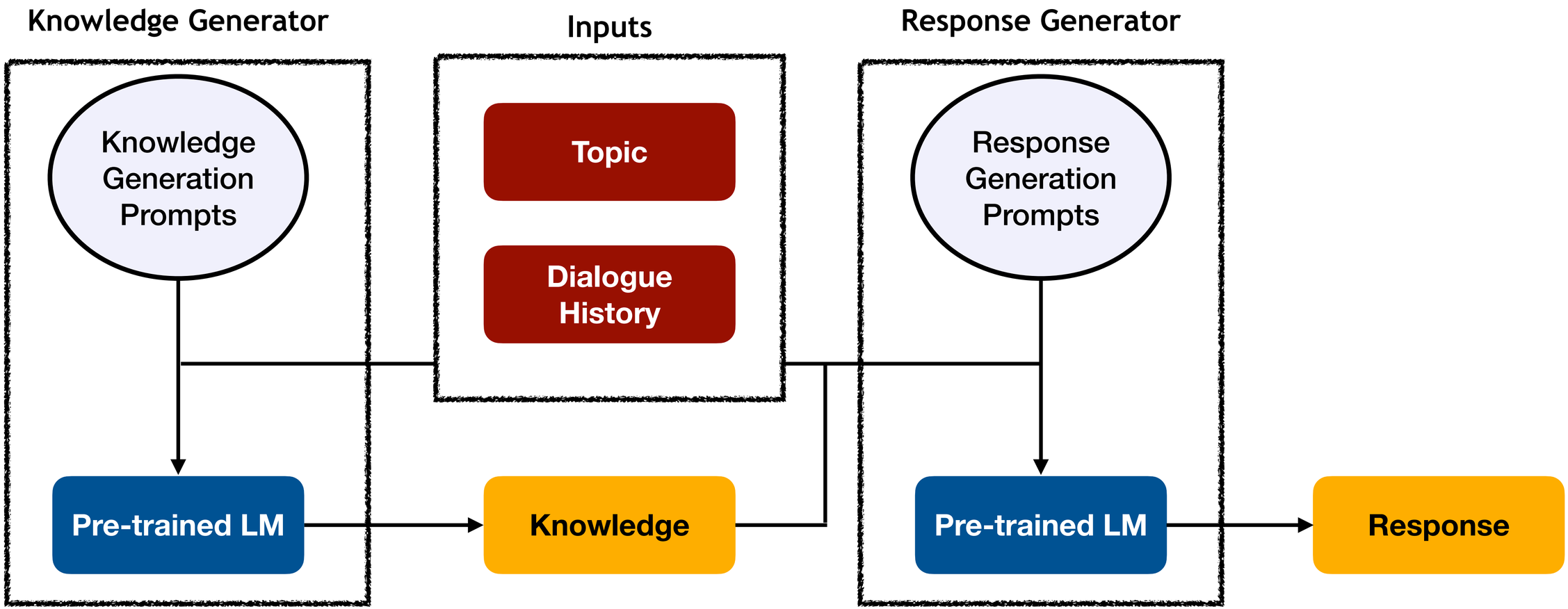
(Liu et al, ACL Findings 2022)

Multi-Stage Dialogue Prompting Framework (MSDP)



(Liu et al, ACL Findings 2022)

Multi-Stage Dialogue Prompting Framework (MSDP)



(Liu et al, ACL Findings 2022)

Knowledge Generator

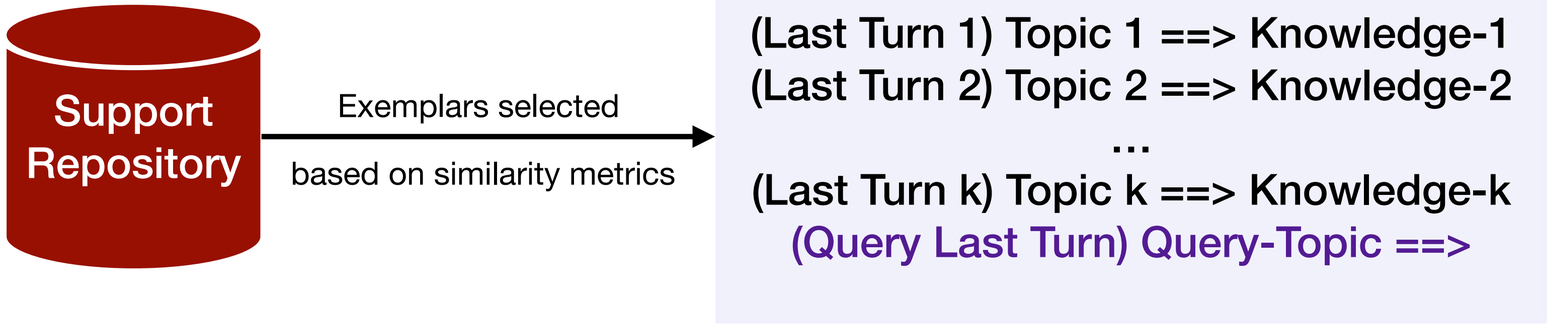
Knowledge Generator

Typically Train sets!

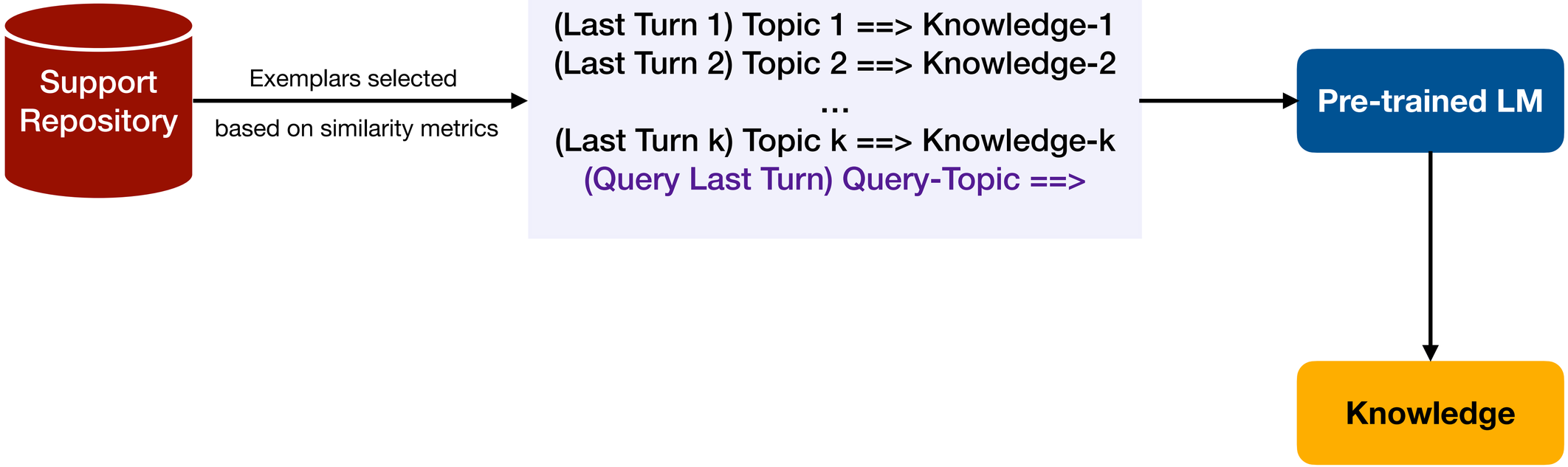
Topic

Dialogue History

Knowledge Generator



Knowledge Generator



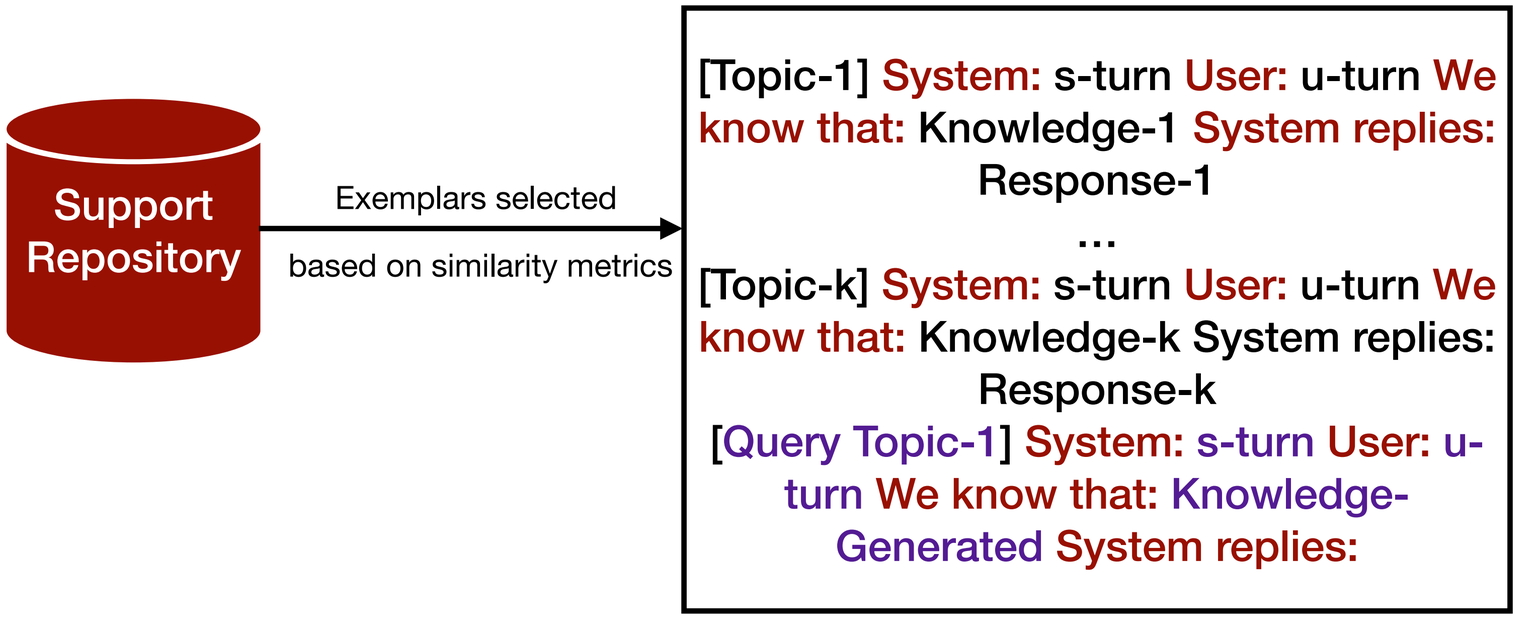
Response Generator

Response Generator

Topic

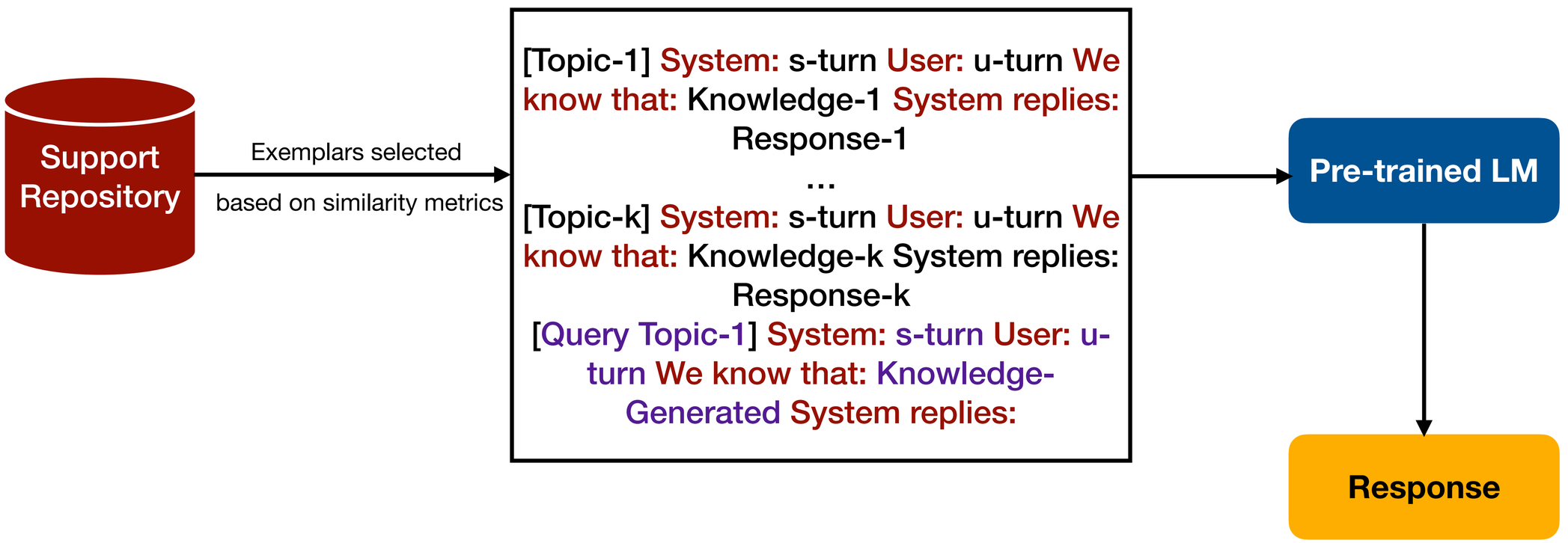
Dialogue
History

Response Generator



Knowledge

Response Generator

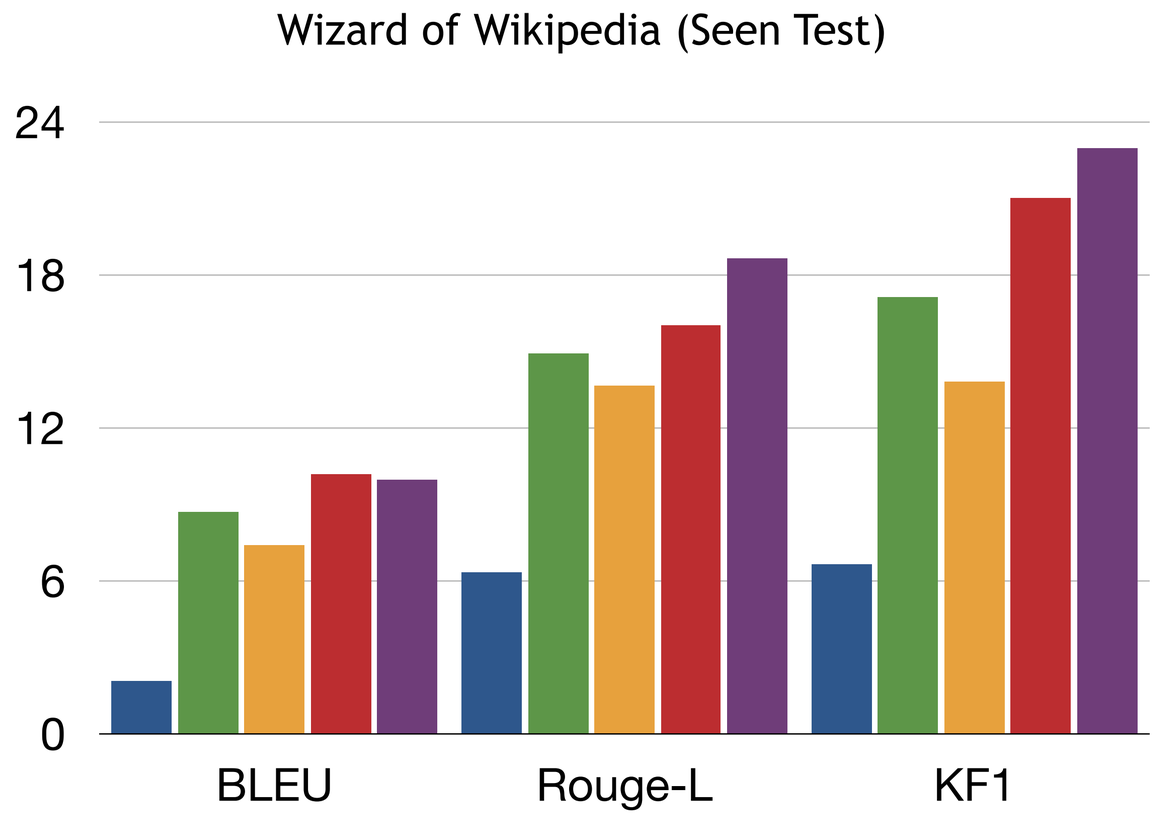


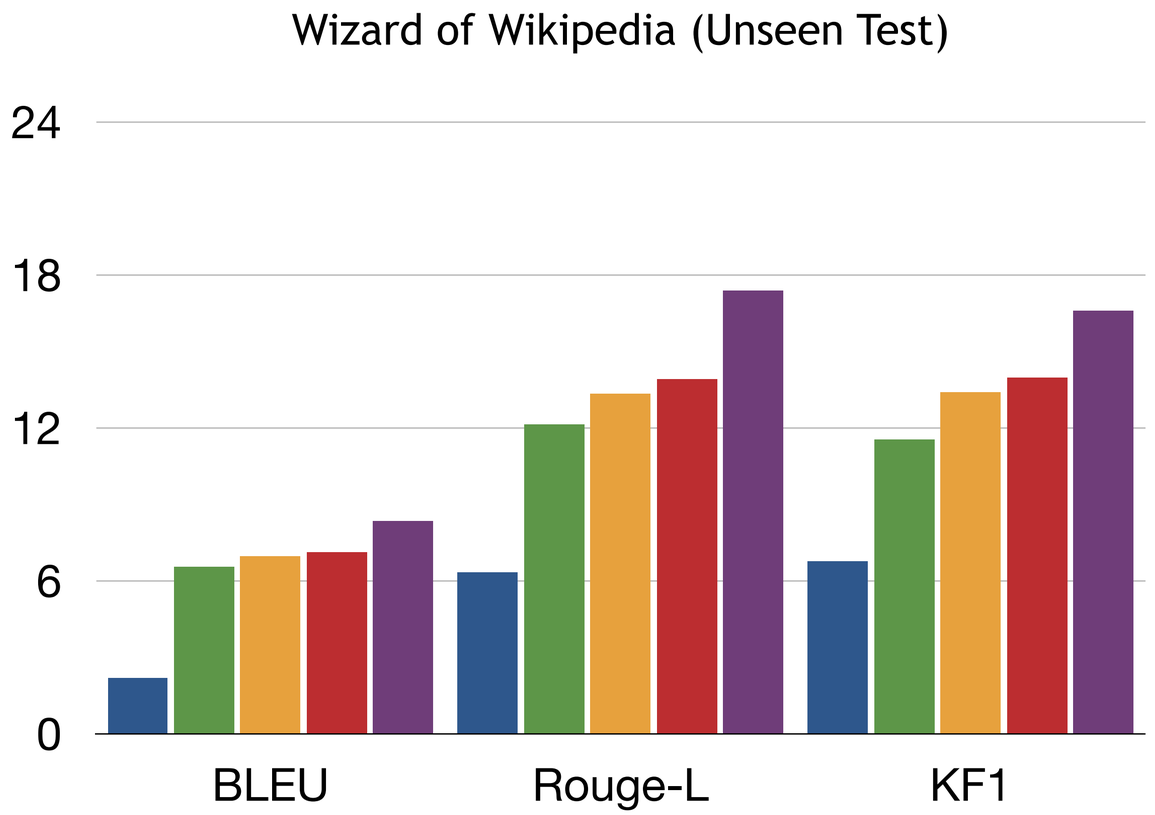
Results

PPLM [Dathathri et al., 2019] FCM w/ DPR (seen) FCM w/ DPR (Wiki) FCM w/ MSDP-KG MSDP

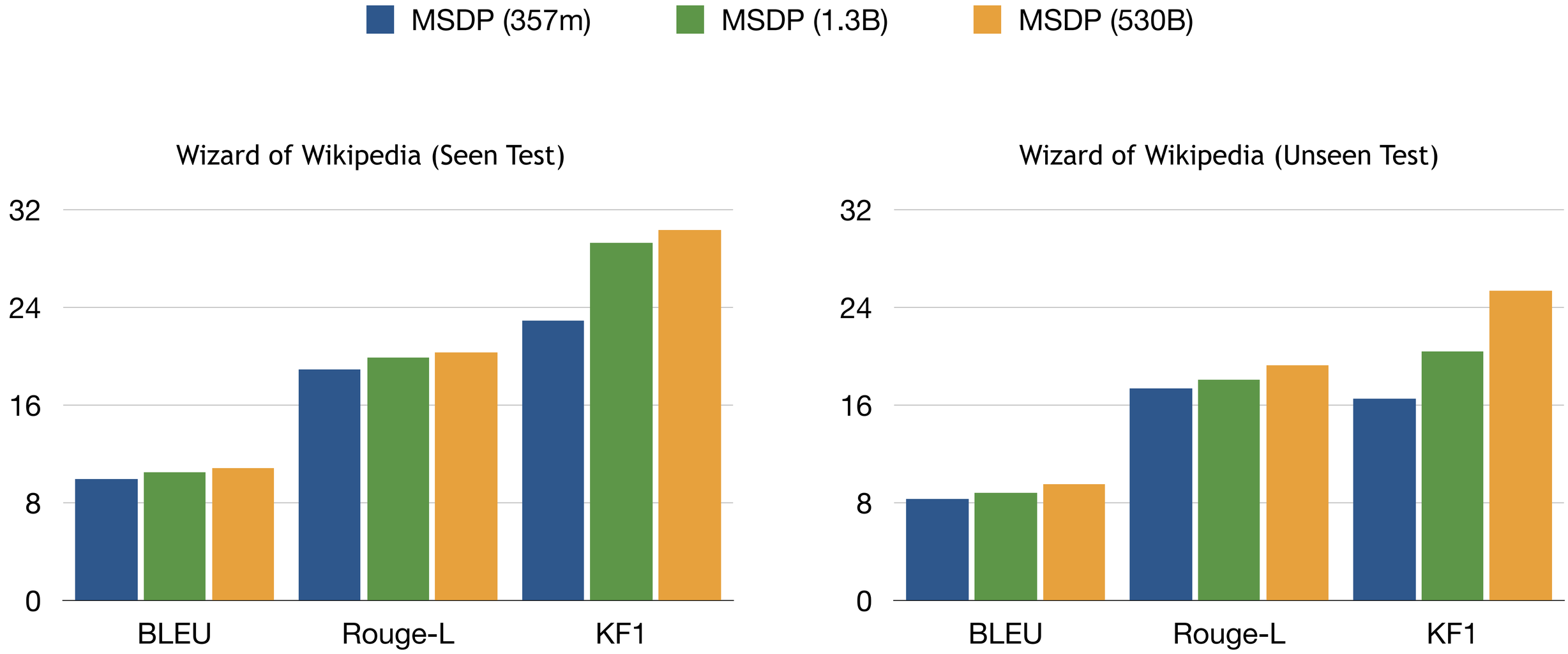
FCM - Fine-tuned Conversation Model

DPR - Dense Passage Retrieval [Karpukhin et al., 2020] fine-tuned on WoW





Model Size and Response Generation



Summary

- The retriever is limited to the support repository set
  - Can retrieve information that is not relevant to dialogue context
- Model generates information relevant to the dialogue context
- More details in the paper
  - Knowledge generation
  - Human Evaluation
  - Ablation on choosing the type on encoder for semantic similarity
  - Ablation on changing the size of k (number of shots)