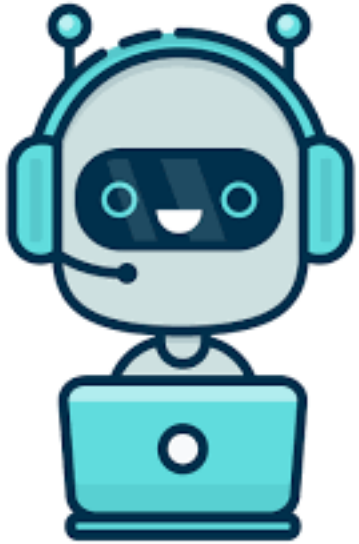
Controllable Text Generation

Should machines reflect the way humans interact in society?

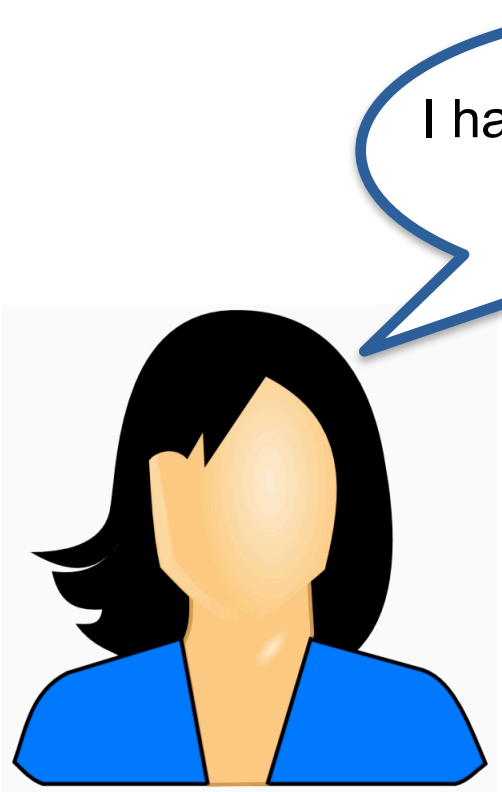
Shrimai Prabhumoye

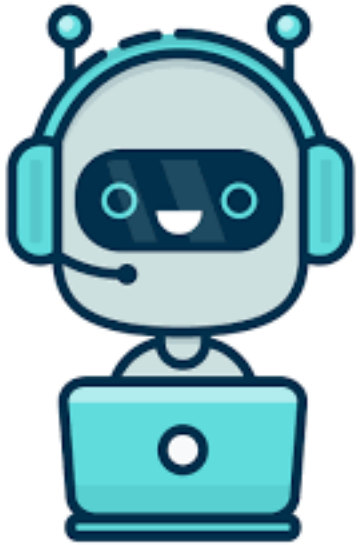
Carnegie Mellon University
Language Technologies Institute

Current Systems don’t control Style

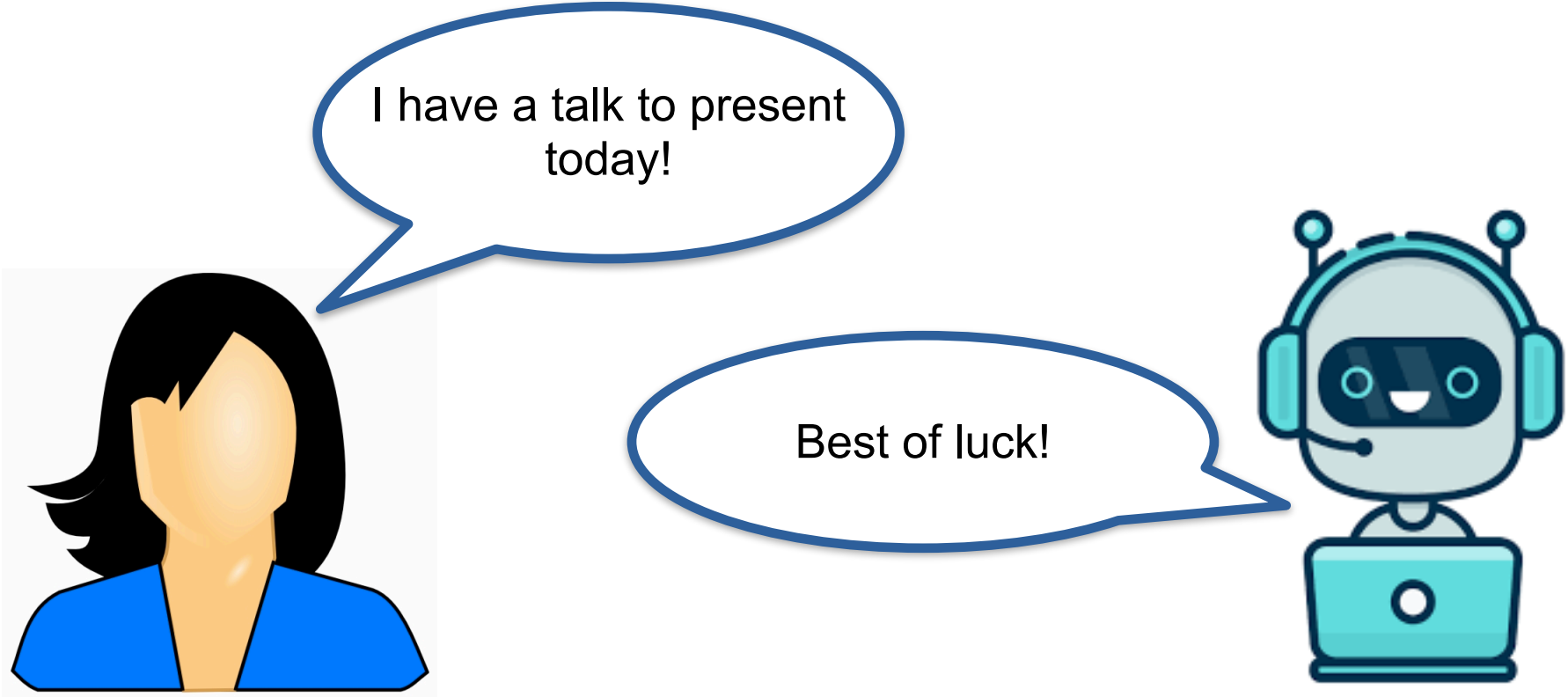


Current Systems don’t control Style



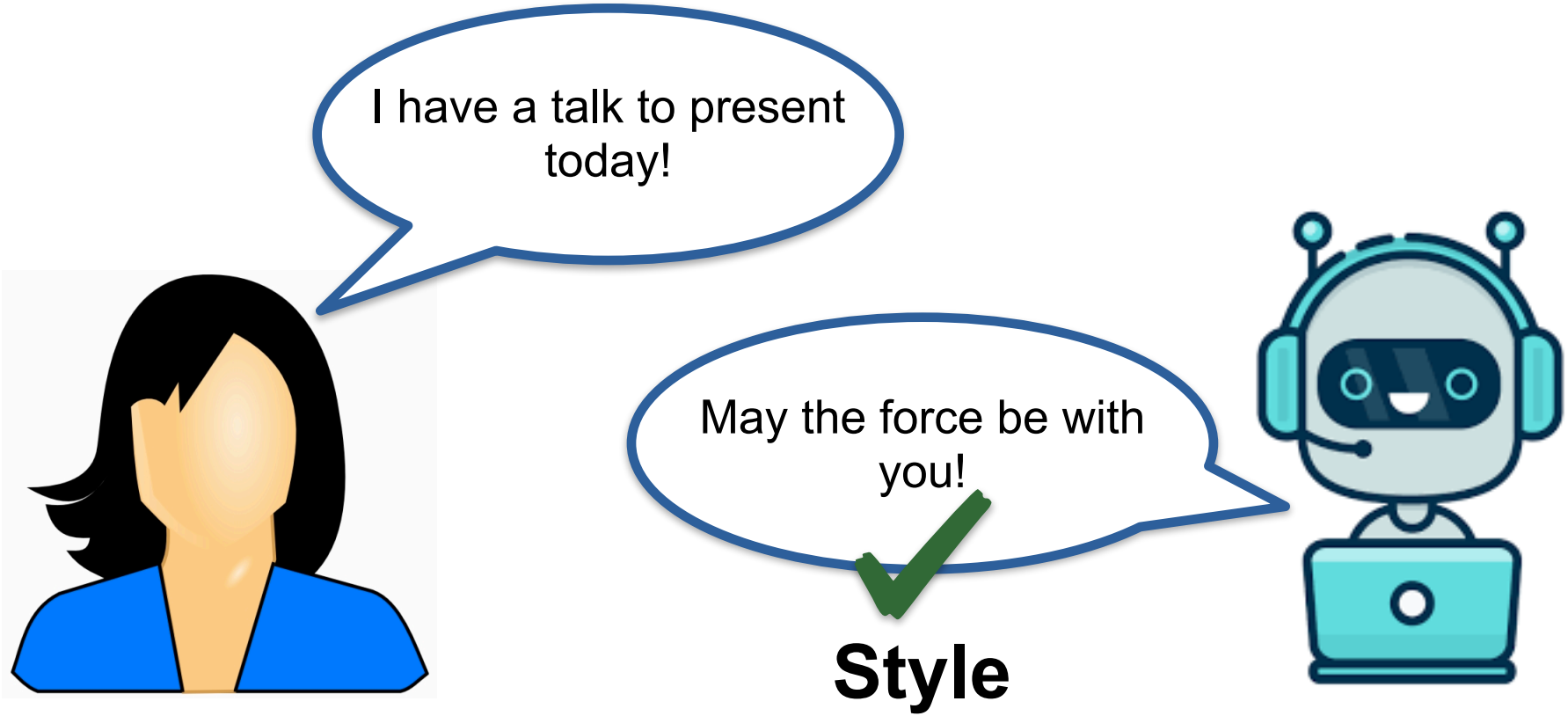


Current Systems don’t control Style

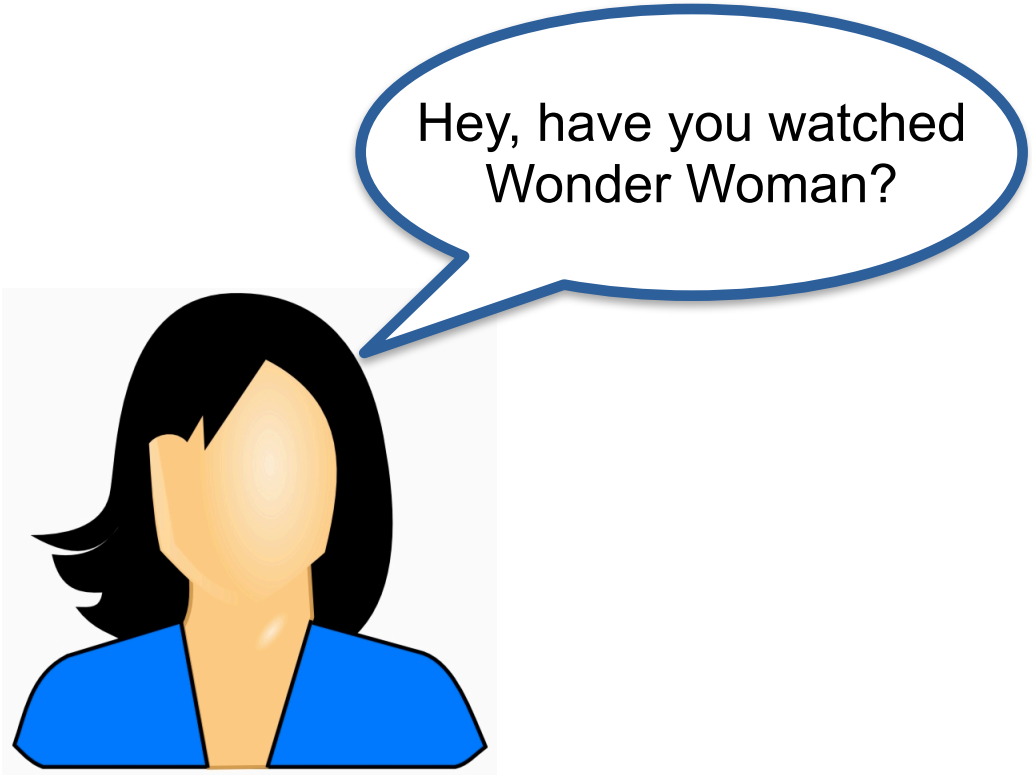


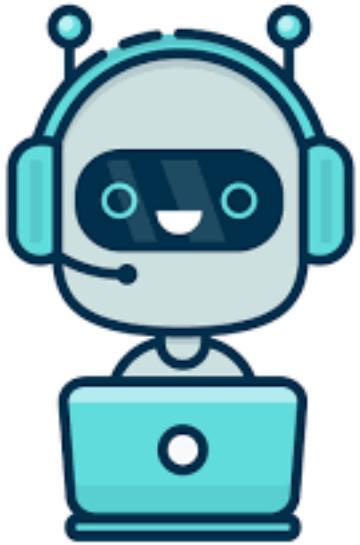
Current Systems don’t control Style

Current Systems don’t control Style

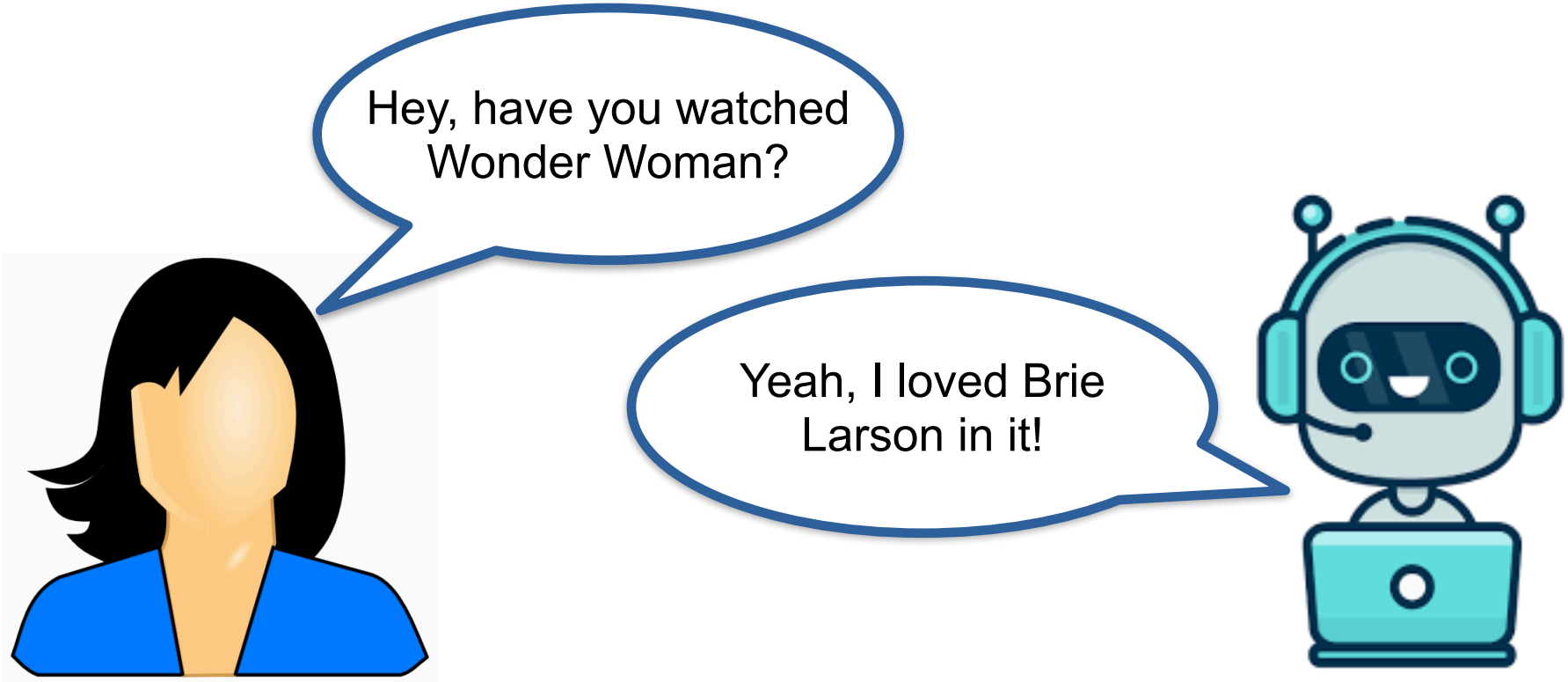


Current Systems don’t control Content

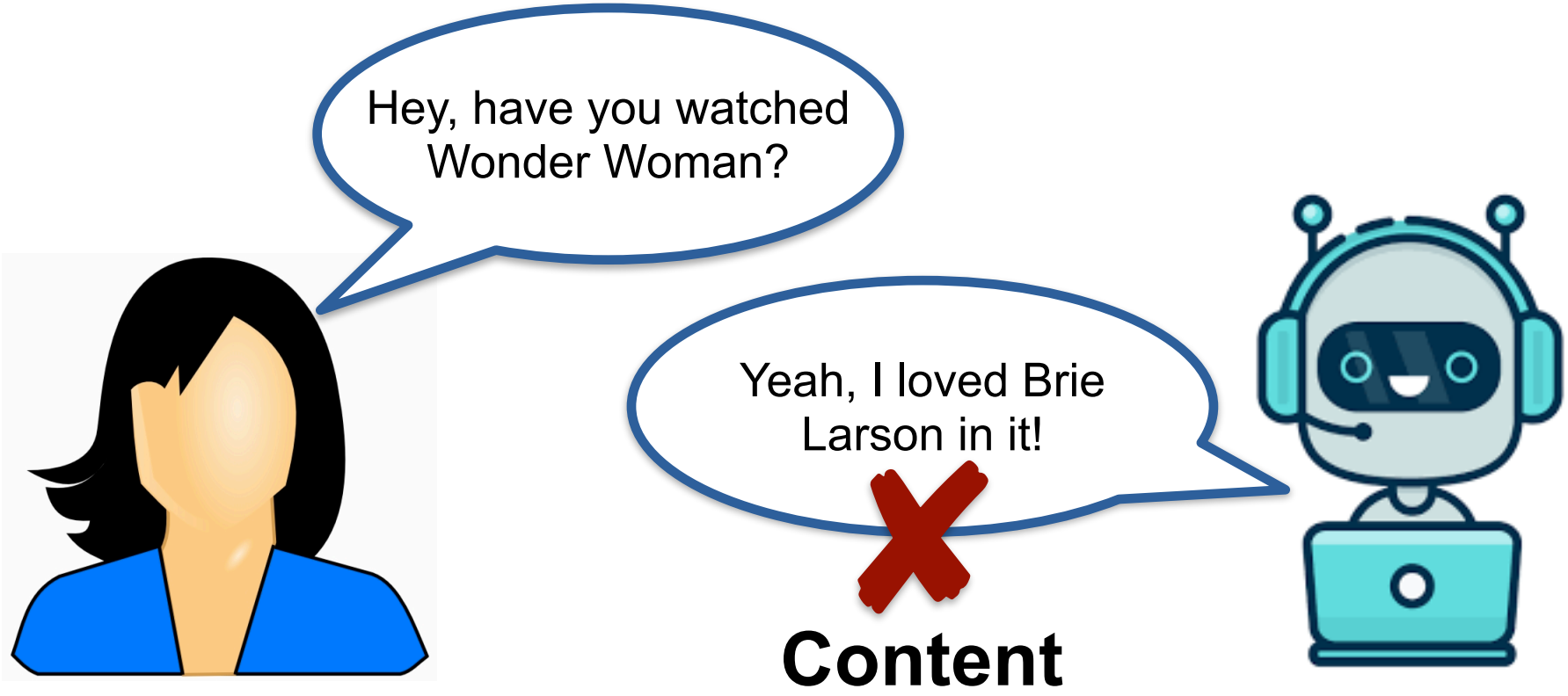




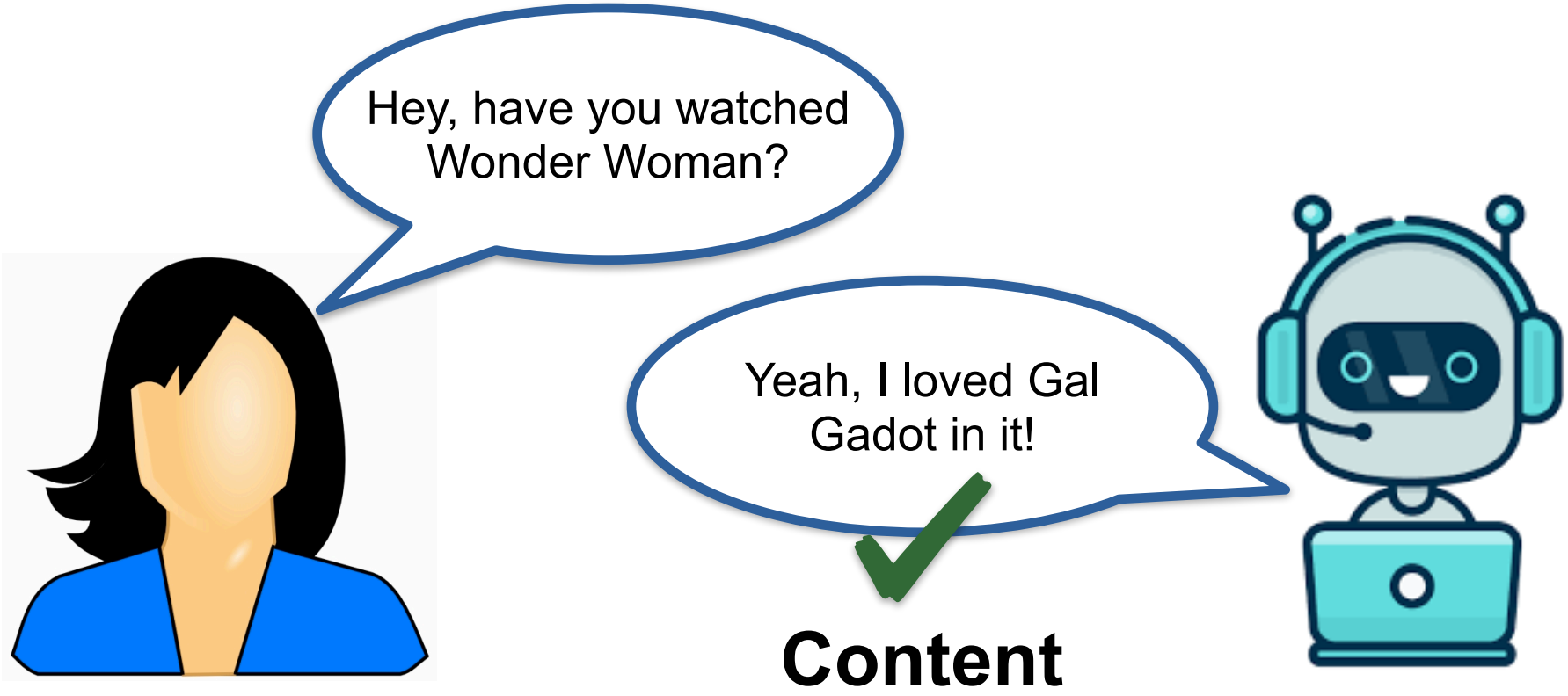
Current Systems don’t control Content



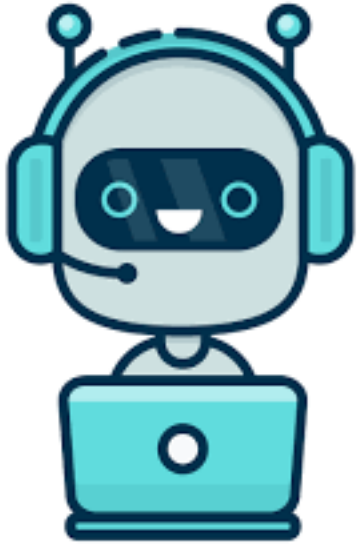
Current Systems don’t control Content



Current Systems don’t control Content

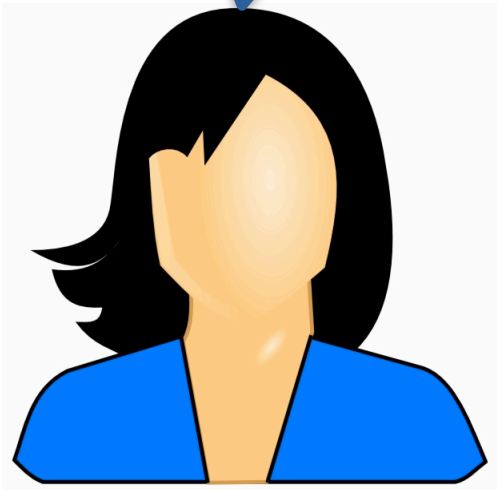


Current Systems don’t control Structure



Current Systems don’t control Structure

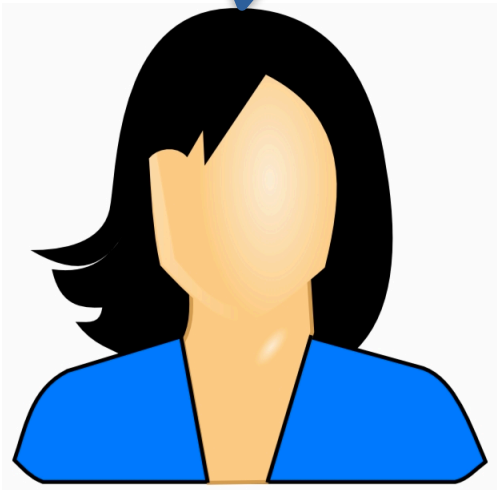
Tell me the story of Avengers Endgame!



Yeah, Ironman steals the Infinity Stones back from Thanos
…
Ironman builds a time machine to save the world.

Current Systems don’t control Structure

Tell me the story of Avengers Endgame!



Yeah, Ironman steals the Infinity Stones back from Thanos
…
Ironman builds a time machine to save the world.

Structure

Current Systems don’t control Structure

Tell me the story of Avengers Endgame!

Yeah, Thor decapitates Thanos
…
Ironman steals the Infinity Stones back from Thanos

Structure

GPT-2 cannot control style and content

Write With Transformer gpt2 ⓘ

Shuffle initial text Trigger autocomplete or tab Select suggestion ↑ ↓ and enter

Can you tell me the story of Avengers Endgame? Did you get a chance to play it when it was released?

- Mitch: It
- I know you are very busy
- I got to play

Who was your favorite in Tenet? I really liked the way he was able to

- get people to care and really get
- pull off a game plan.
- get away with things, the way

Can you tell me the different ways of saying 'Please be quiet.'? How many ways ? What's the difference

- between
- ?

Can you tell me the story of Tenet? How did he get his job as the director

- of the CIA?↵↵↵He
- of the CIA in 1997 when his
- of the CIA? Did he

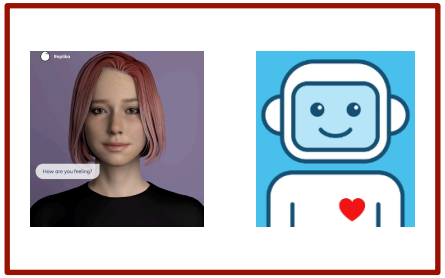
NLP Applications are on the rise!

NLP Applications are on the rise!

Personal Assistants

NLP Applications are on the rise!

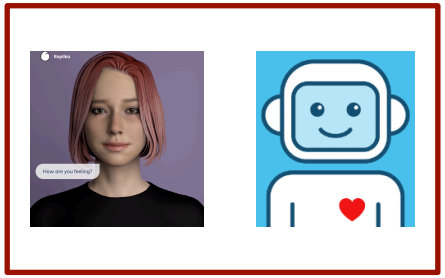
Personal Assistants



Empathy/Compassion bots

NLP Applications are on the rise!

Personal Assistants

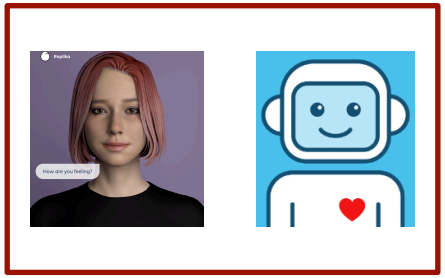


Empathy/Compassion bots

Call-center/FAQ bots

NLP Applications are on the rise!

Personal Assistants



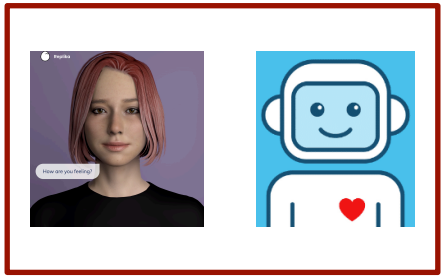
Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing

NLP Applications are on the rise!

Personal Assistants



Empathy/Compassion bots

Call-center/FAQ bots

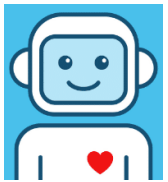
Ubiquitous Computing

Other Applications

NLP Applications are on the rise!

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

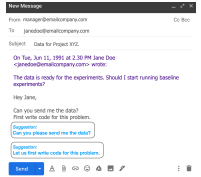
Personal Assistants



Empathy/Compassion bots

Call-center/FAQ bots

Ubiquitous Computing



Other Applications

NLP Applications are on the rise!

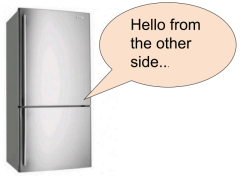
- Style of responses
- Persona of the bot
- Topic of responses

Personal Assistants

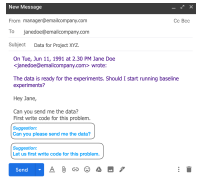
- Emotions (style)
- Sensitivity of content

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

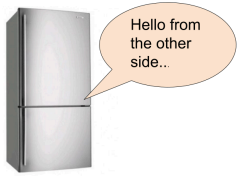
Personal Assistants

- Emotions (style)
- Sensitivity of content

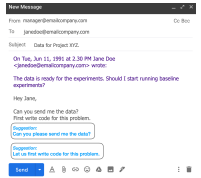
Empathy/Compassion bots

- 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

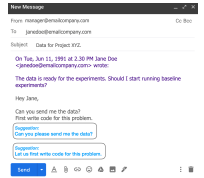
Empathy/Compassion bots

- 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

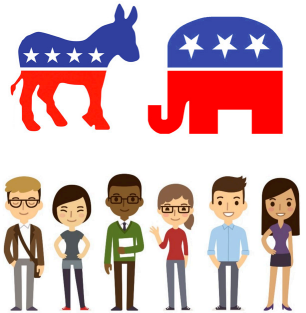
Ubiquitous Computing

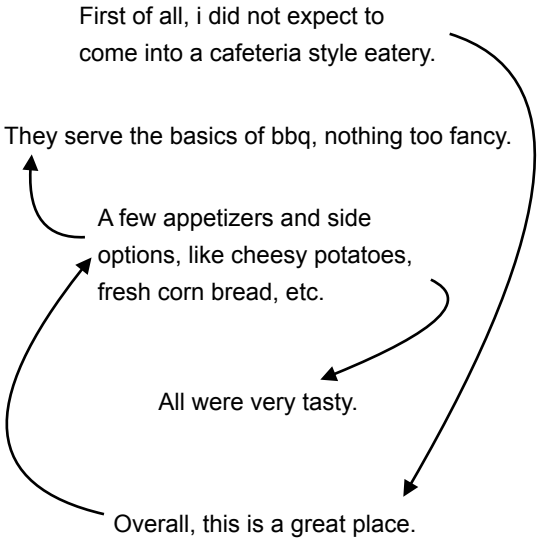
- Plot of stories (structure)
- Sequence of topics in stories (structure)
- Politeness of emails (style)

Other Applications

The key to human-like generation is

The key to human-like generation is





Style

Content

Structure

Style is “how” an intent is expressed

- Expressed in the choice of *words or phrases* as well as *syntactic structures* used to convey information

(Coupland, CUP 2007; Eckert, Language 2019; Hovy, JOP 1987; Hu et al, ICML 2017; Shen et al, NeurIPs 2017; Li et al, NAACL 2018; Prabhumoye et al, ACL 2018; Fu et al, AAAI 2018; John et al, ACL 2019; Sudhakar et al, EMNLP-IJCNLP 2019; Madaan et al, ACL 2020)

“Do you have any code that we can look at?”

“Any code? Or do I have to guess at how you did it?”

(Danescu-Niculescu-Mizil, 2013)

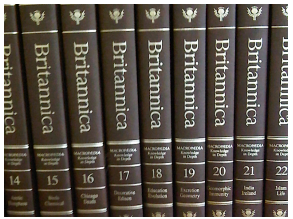
Content is “what” is to be conveyed

- Communication: individuals *exchange information* and influence one another through a common system of symbols and signs.
- Leverage information in the unstructured form

(Liu et al, ICLR 2018; Zhou et al, EMNLP 2018; Ghazviinejad et al, AAAI 2018; Dinan et al, ICLR 2019; Prabhumoye et al, NAACL 2019)

Wikipedia

Gutenberg



Encyclopedia

Internet

Structure is ordering of information

- Communication is well *planned* and structured
- Sentence ordering: *ordering* of information

1 [Obama Birth & 44th president]
Barack Hussein Obama II (born August 4, 1961) … 44th president of the US from 2009 to 2017.

2 [Democrat, first African-American president]
A member of the Democratic party, Obama was the first African-American president of the US.

3 [Previously an Illinois senator]
He previously served as a US senator from Illinois from 2005 to 2008 …

3 [Previously an Illinois senator]
He previously served as a US senator from Illinois from 2005 to 2008 …

2 [Democrat, first African-American president]
A member of the Democratic party, Obama was the first African-American president of the US.

1 [Obama Birth & 44th president]
Barack Hussein Obama II (born August 4, 1961) … 44th president of the US from 2009 to 2017.

(Barzilay and Lee, NAACL 2004; Barzilay and Lapata, CL 2008; Li and Hovy, EMNLP 2014; Cui et al, EMNLP 2018; Prabhumoye et al, NAACL 2020)

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

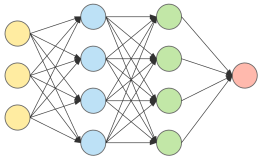
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

Style

ACL '20
Storytelling '19
ACL '18

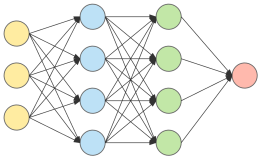
Content

NAACL '21
NAACL '19
EMNLP '18

Structure

ACL '20

Control Attributes



Modeling

COLING '20

Ethics

NAACL '21
chap. in L.L.C '21
WiNLP '19

Common Themes

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

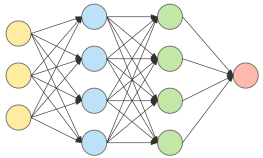
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

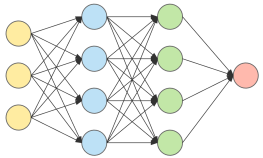
COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

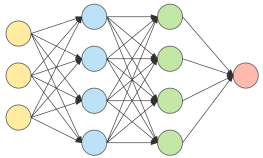


COLING '20

Modeling

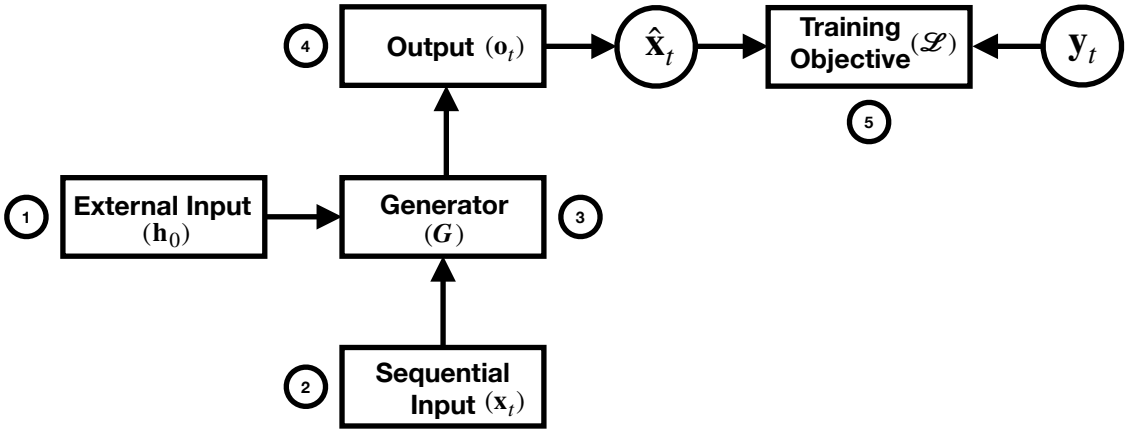
Overview

- Schema of five modules
- Organize prior work
- Understand techniques to add control



COLING '20

Modeling



Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

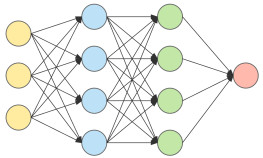
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

- Constraint Solving Problem vs Sequence Prediction Task
- Learn constraints using BERT
- Topological sort to find an order given constraints

Structure

ACL '20

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

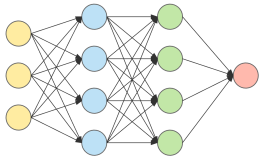
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

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.

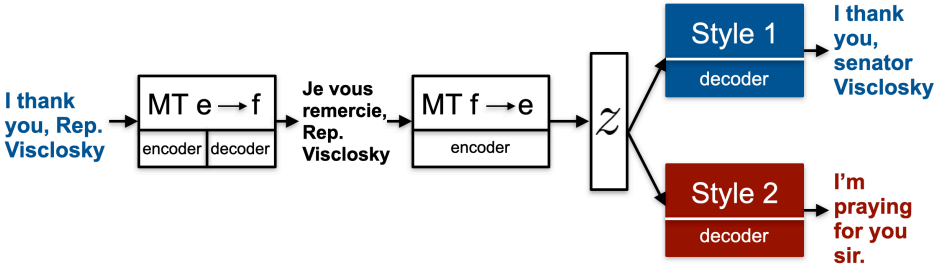
Challenges in style transfer

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

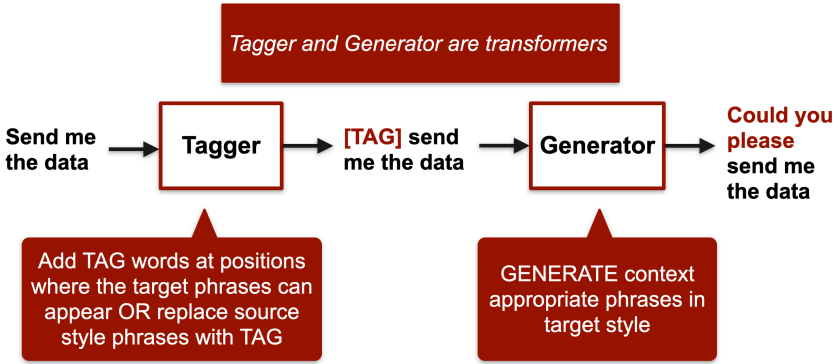
Challenges in style transfer

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

Style Transfer



Back-Translation Style Transfer [ACL, 2018]



Tag and Generate [ACL, 2020]

- Two novel approaches for style transfer
- Political slant transfer and Politeness transfer

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

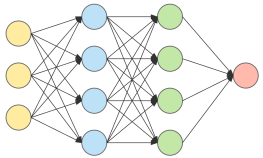
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

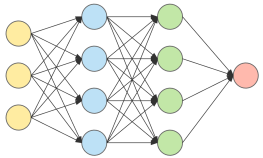
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

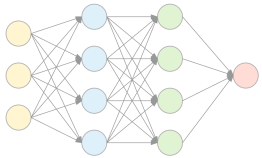
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

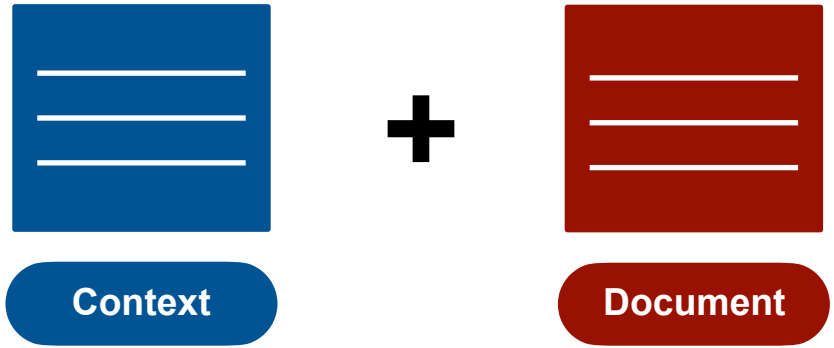
Common Themes

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

(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
Special pages

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.

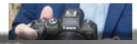
Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

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
Special pages

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

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]

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.

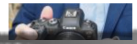
Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

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
Special pages

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.

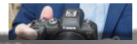
Speaking to AP, David explained that his camera had been mounted on a tripod when the primates

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]

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!

Document

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.

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***!

Article Talk

WIKIPEDIA
The Free Encyclopedia

Home Alone

…pedia, the free encyclopedia

…rican comedy film written and produced by John …s Columbus. It stars Macaulay Culkin, Joe Pesci, …nd Catherine O'Hara. Culkin stars as eight-year-old …e from burglars Harry and Marv (Pesci and Stern) …eaves him behind on their vacation.

…ween February and May 1990 in locations …ginally set to be distributed by Warner Bros., but …Fox when Hughes exceeded Warner's $10 million

Document

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.***

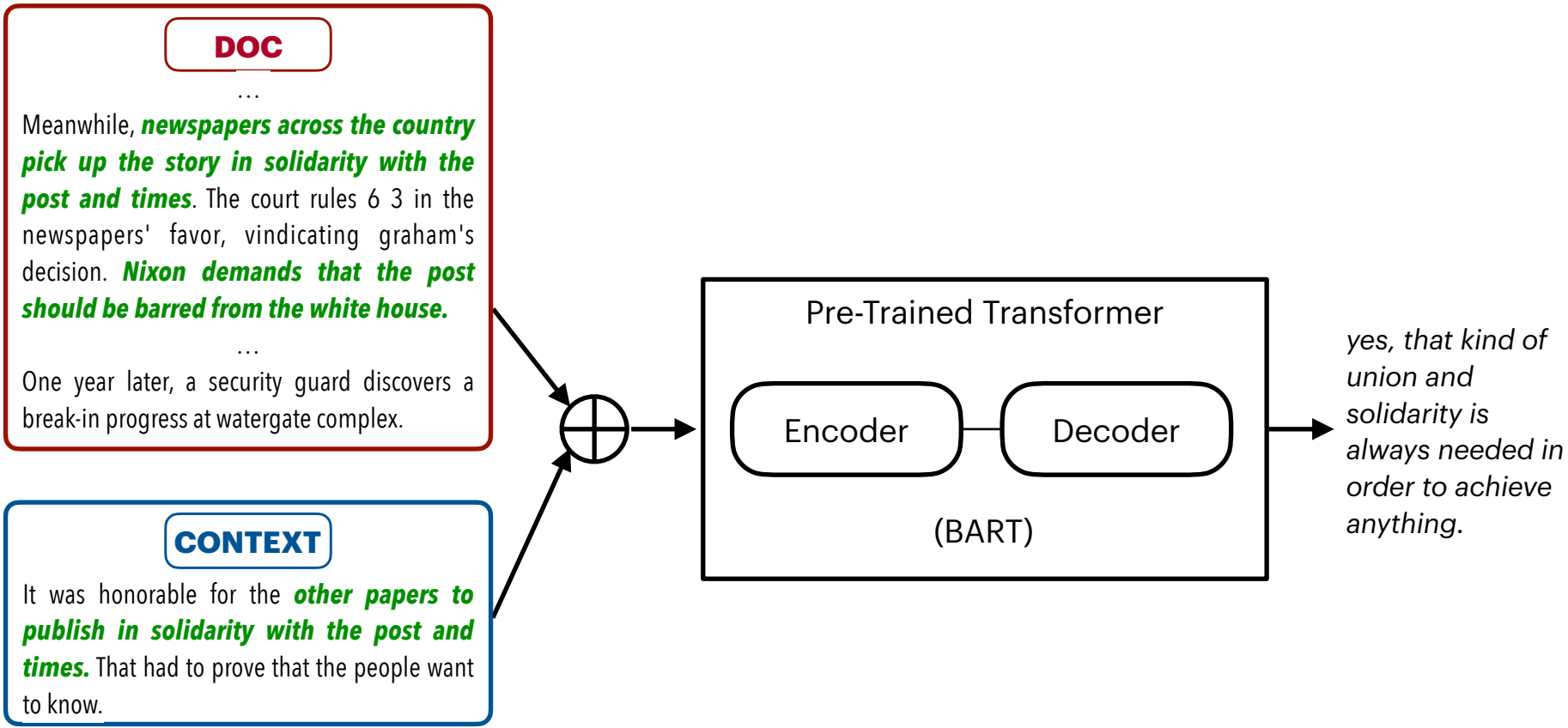
…

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



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

…

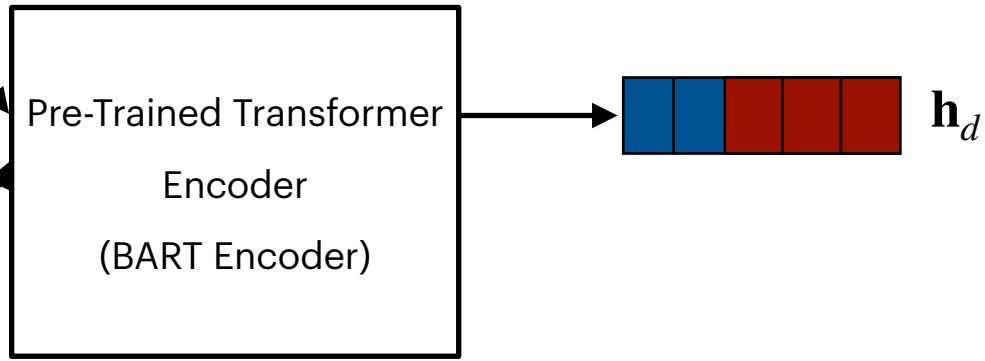
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

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.



Context Driven Representation (CoDR)

DOC

…

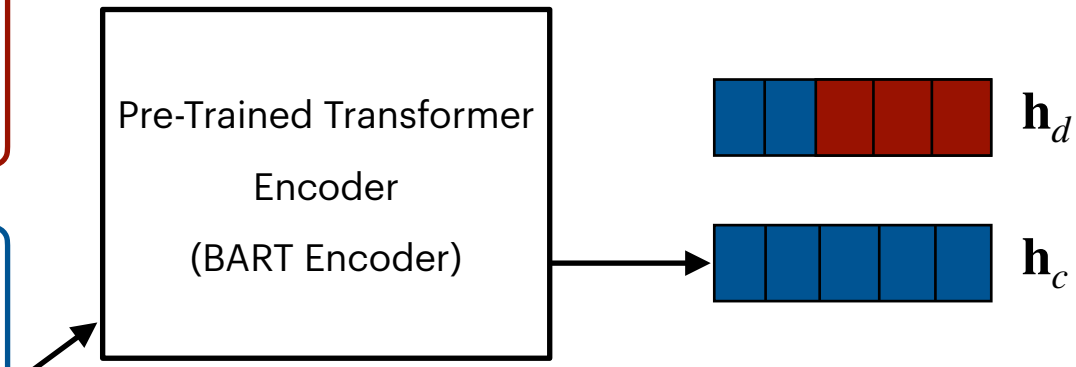
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

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

…

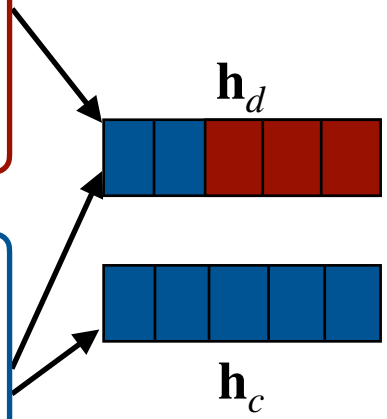
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

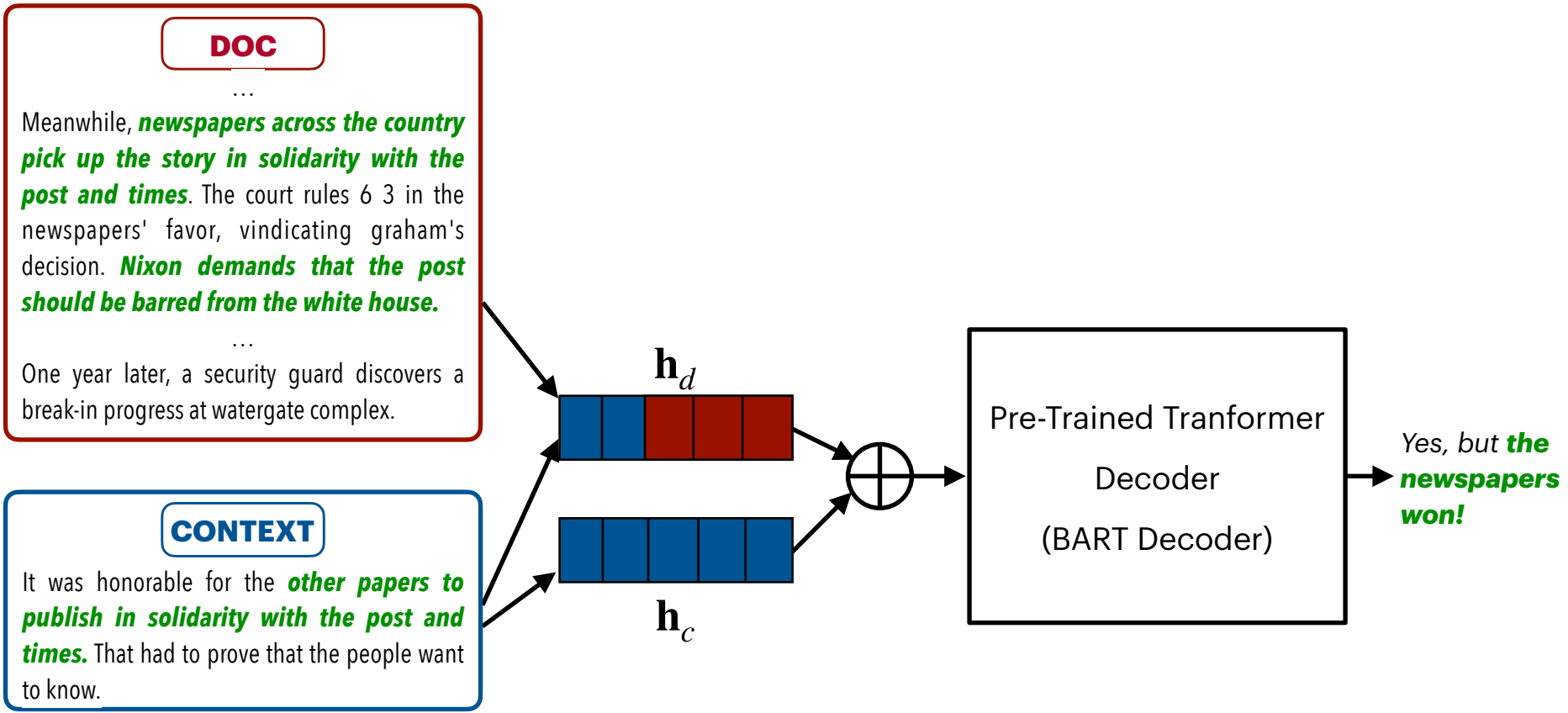
One year later, a security guard discovers a break-in progress at watergate complex.

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)



Document Headed Attention (DoHA)

DOC

…

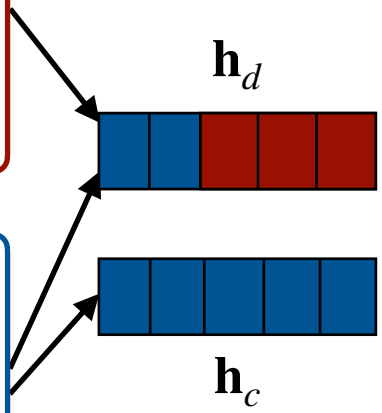
Meanwhile, ***newspapers across the country pick up the story in solidarity with the post and times***. The court rules 6 3 in the newspapers' favor, vindicating graham's decision. ***Nixon demands that the post should be barred from the white house.***

…

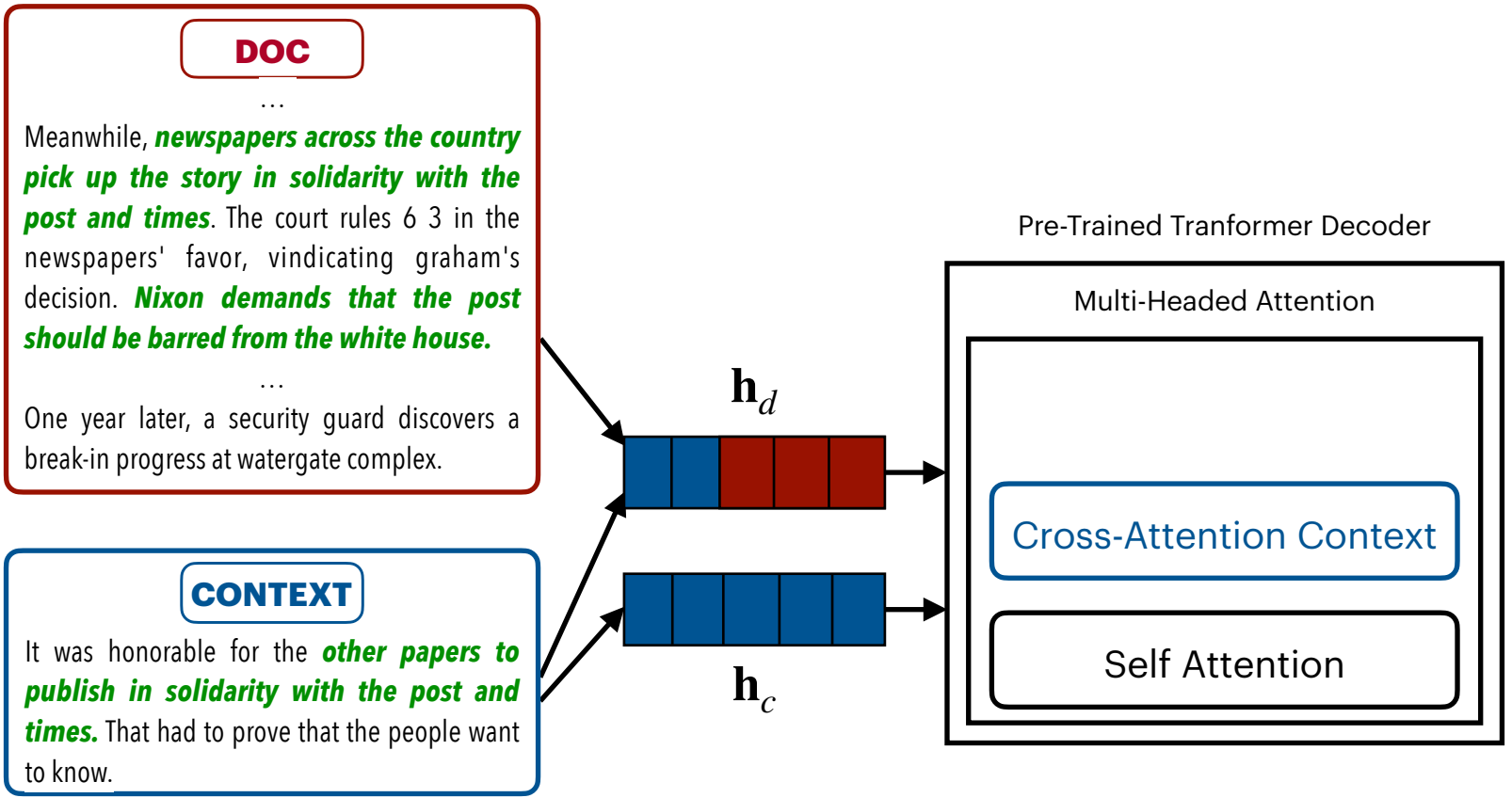
One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

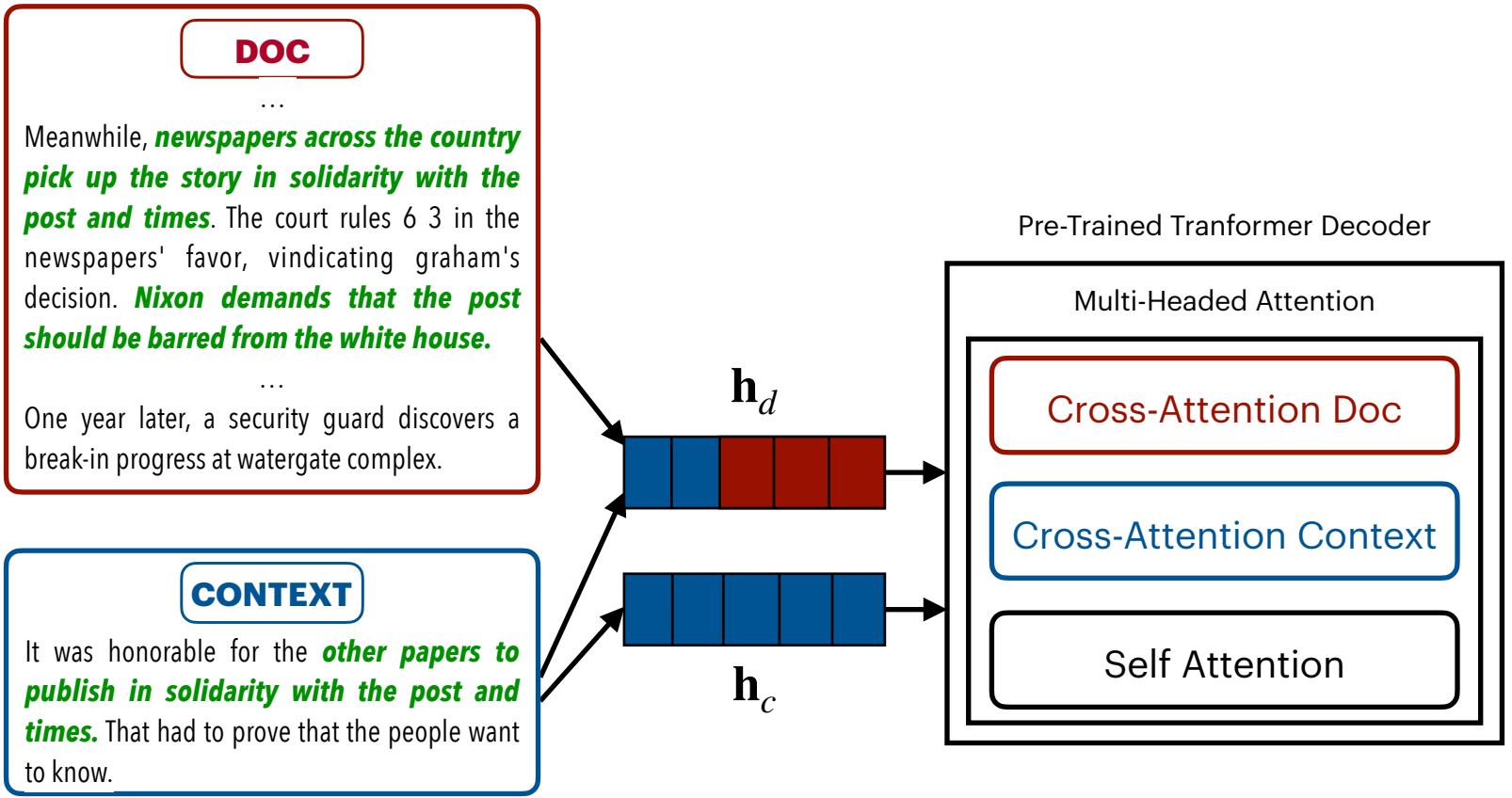
It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.



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.***

…

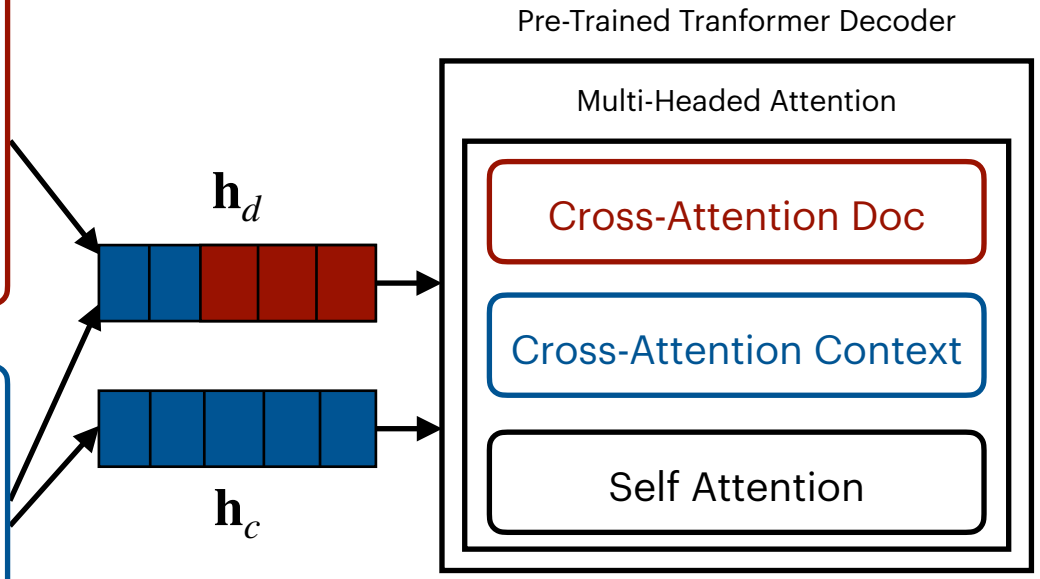
One year later, a security guard discovers a break-in progress at watergate complex.

CONTEXT

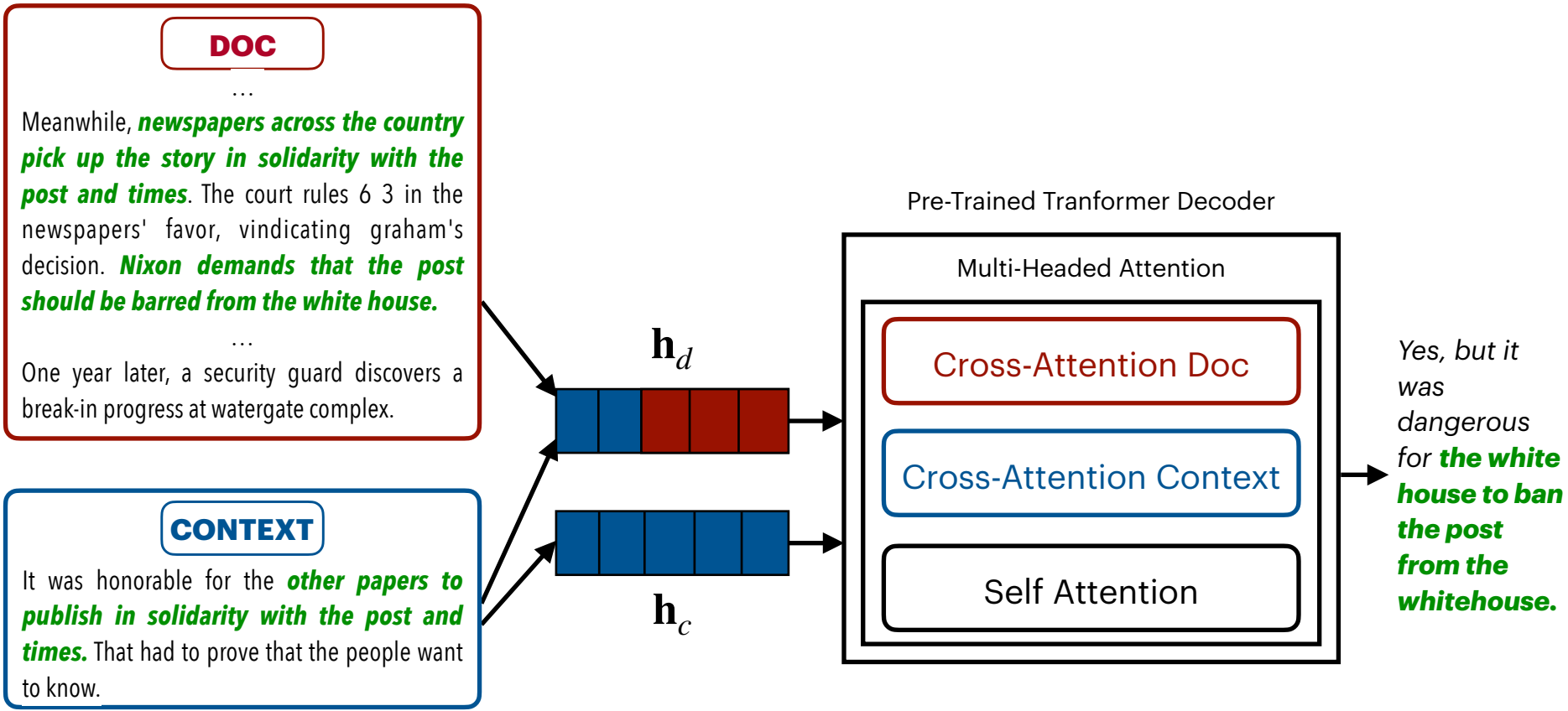
It was honorable for the ***other papers to publish in solidarity with the post and times.*** That had to prove that the people want to know.

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

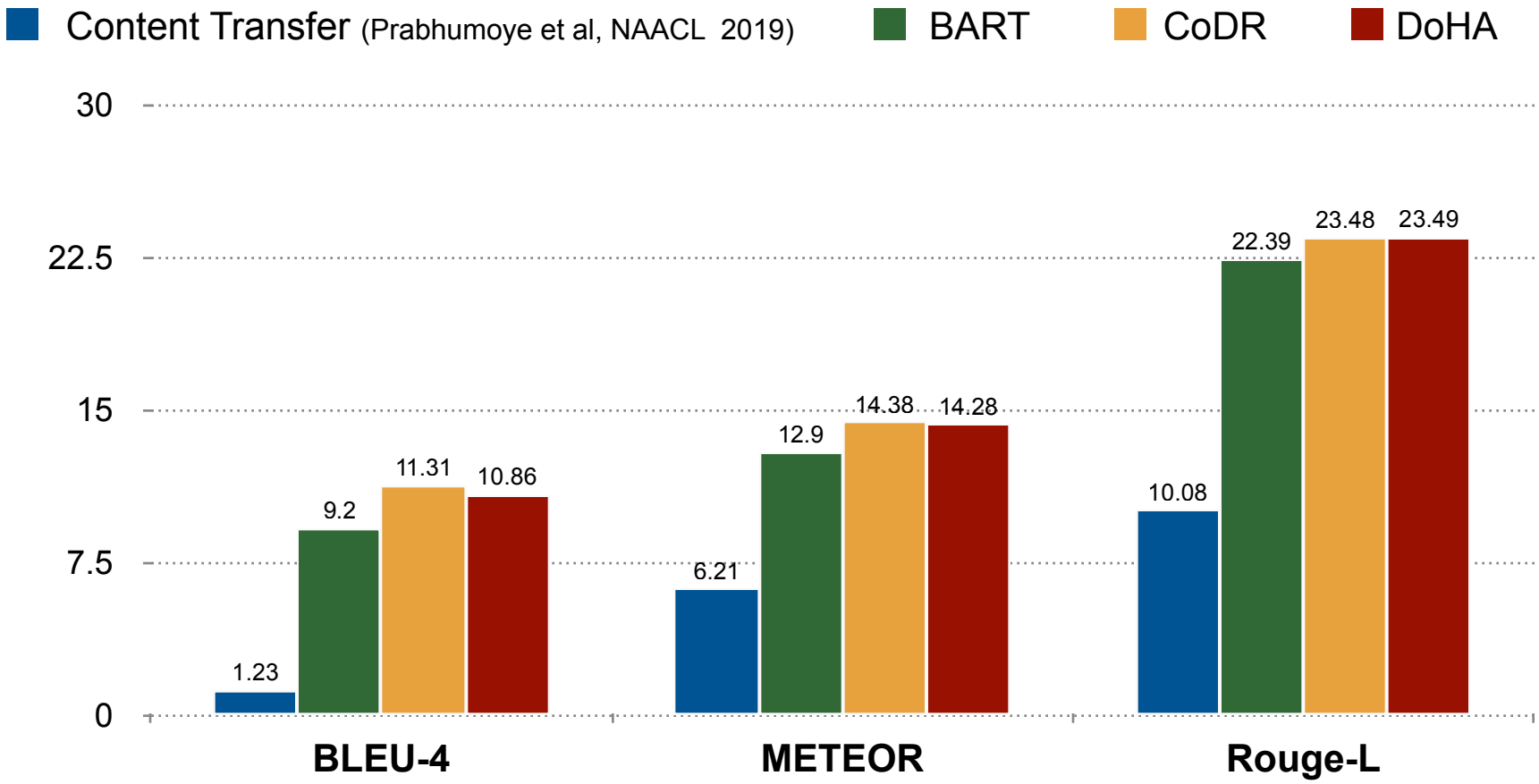
$$\mathbf{H}_i = \texttt{Attention}(Q\mathbf{W}_i^{dQ}, K\mathbf{W}_i^{dK}, V\mathbf{W}_i^{dV})$$



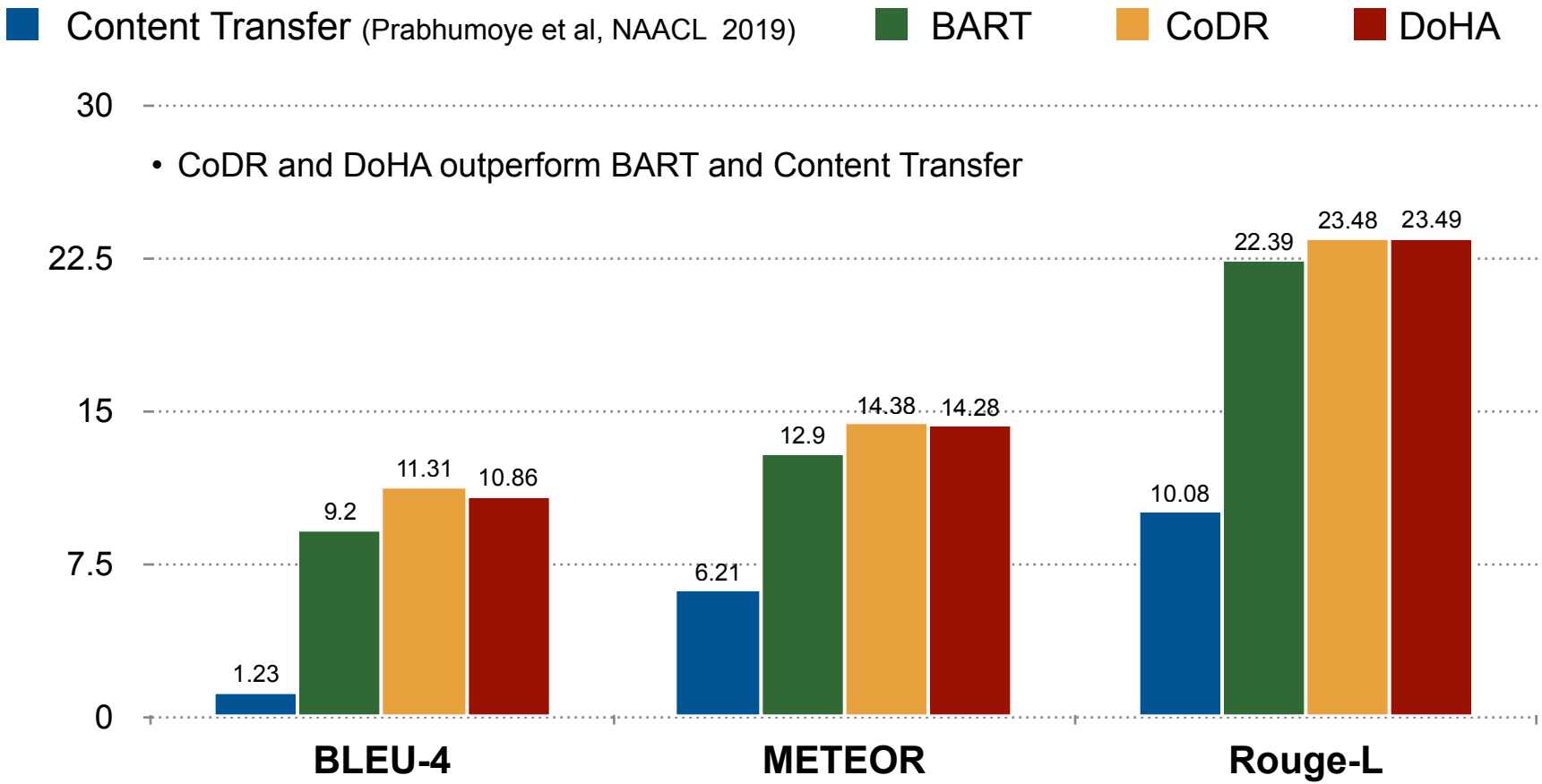
Document Headed Attention (DoHA)



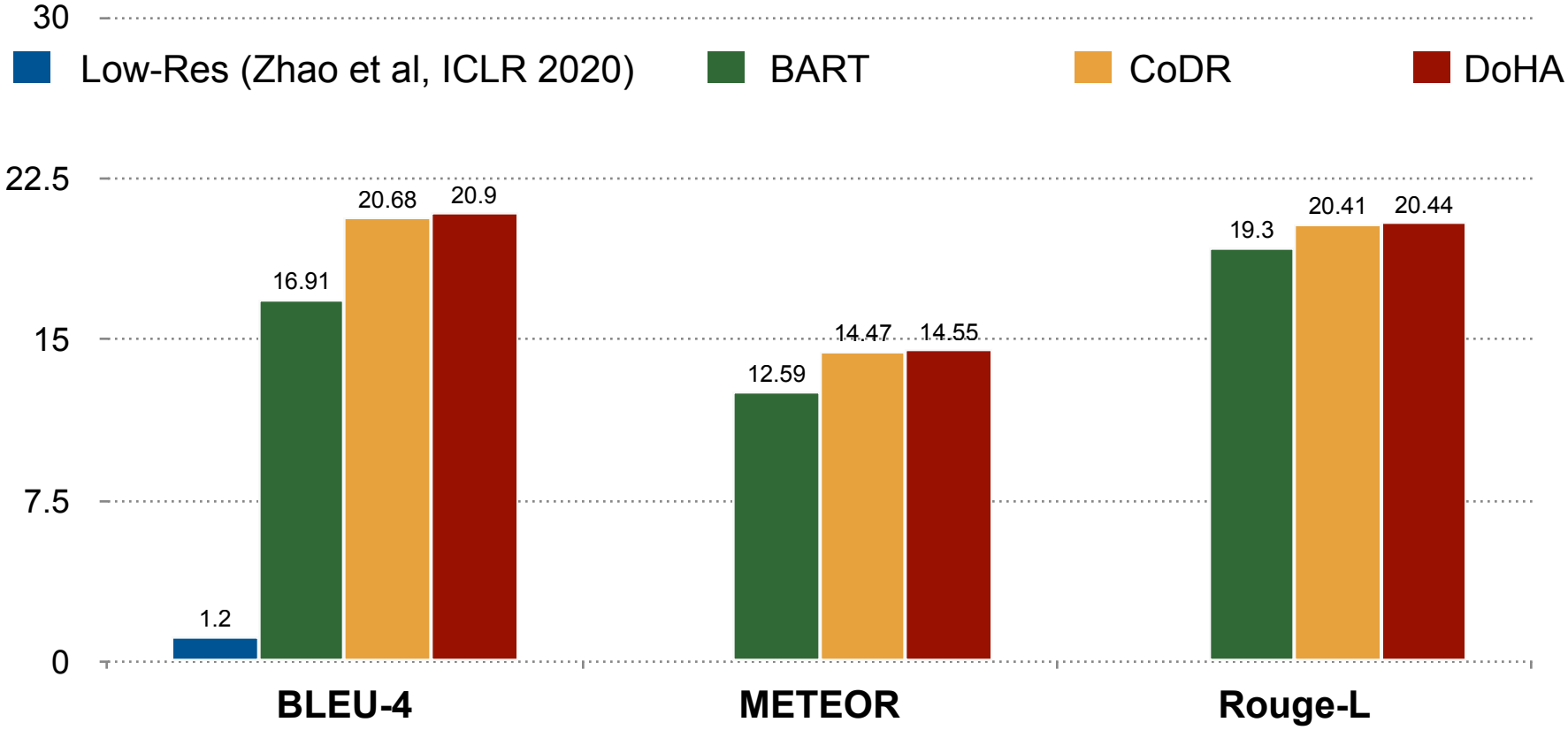
Wikipedia Update Generation



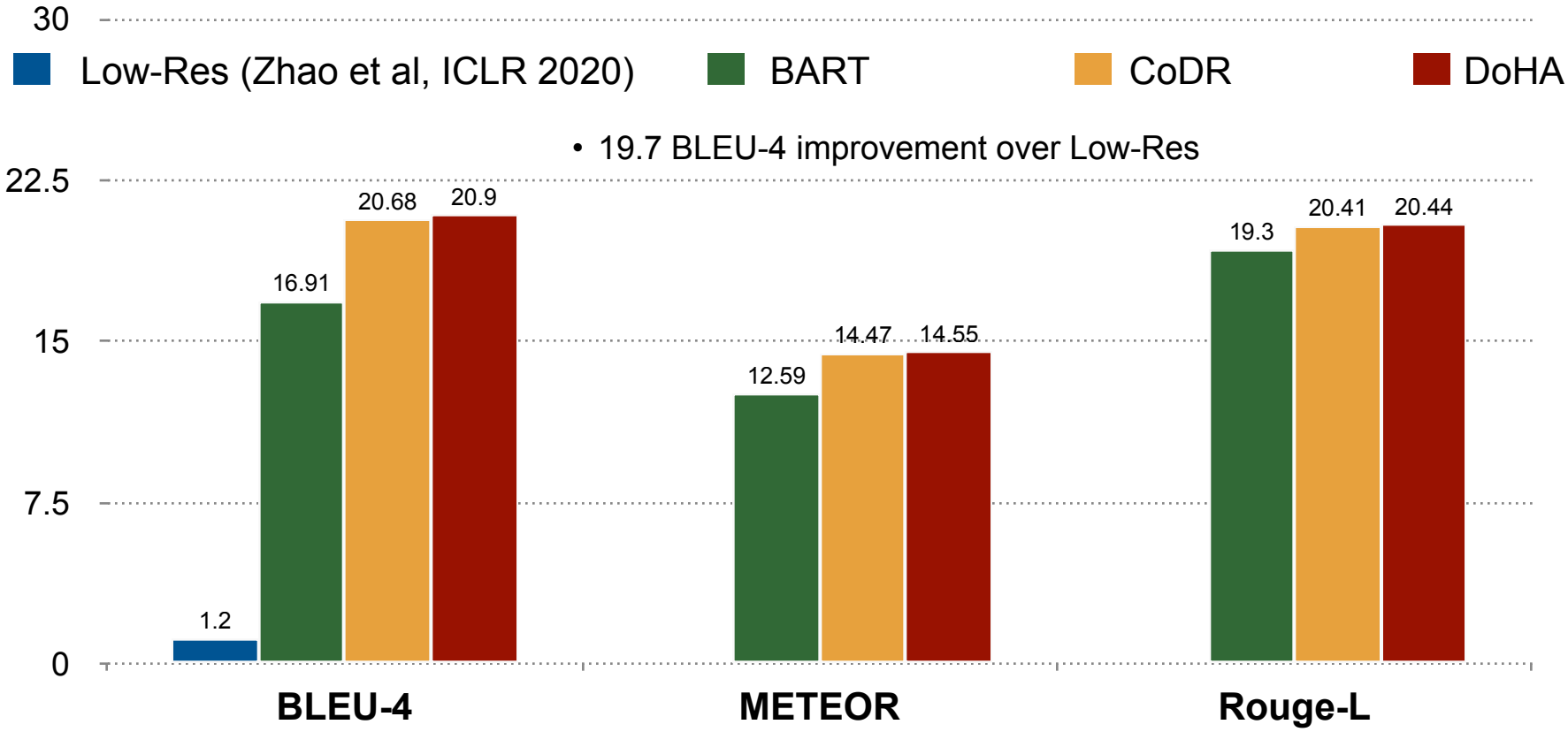
Wikipedia Update Generation



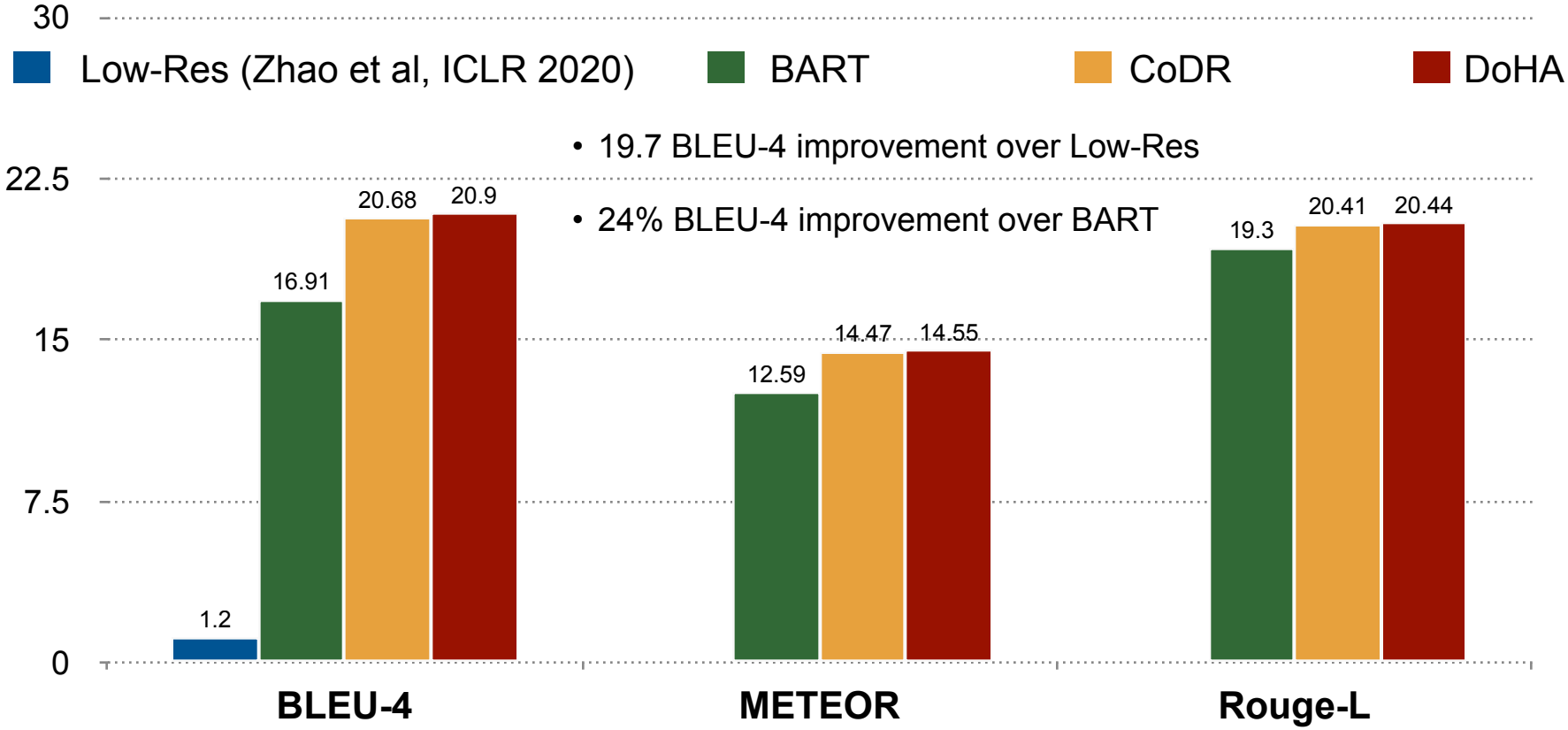
CMU_DoG


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

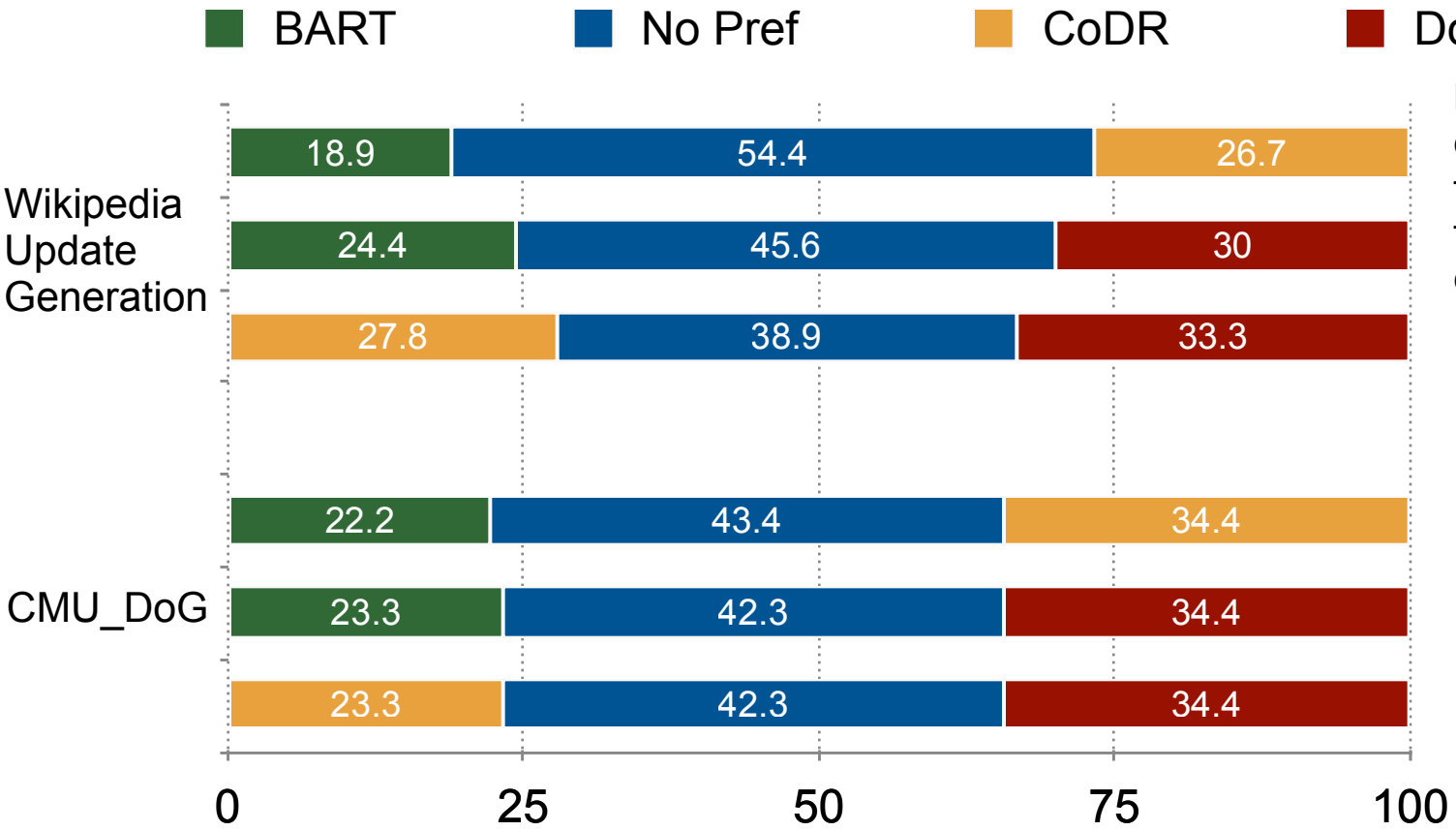
CMU_DoG


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

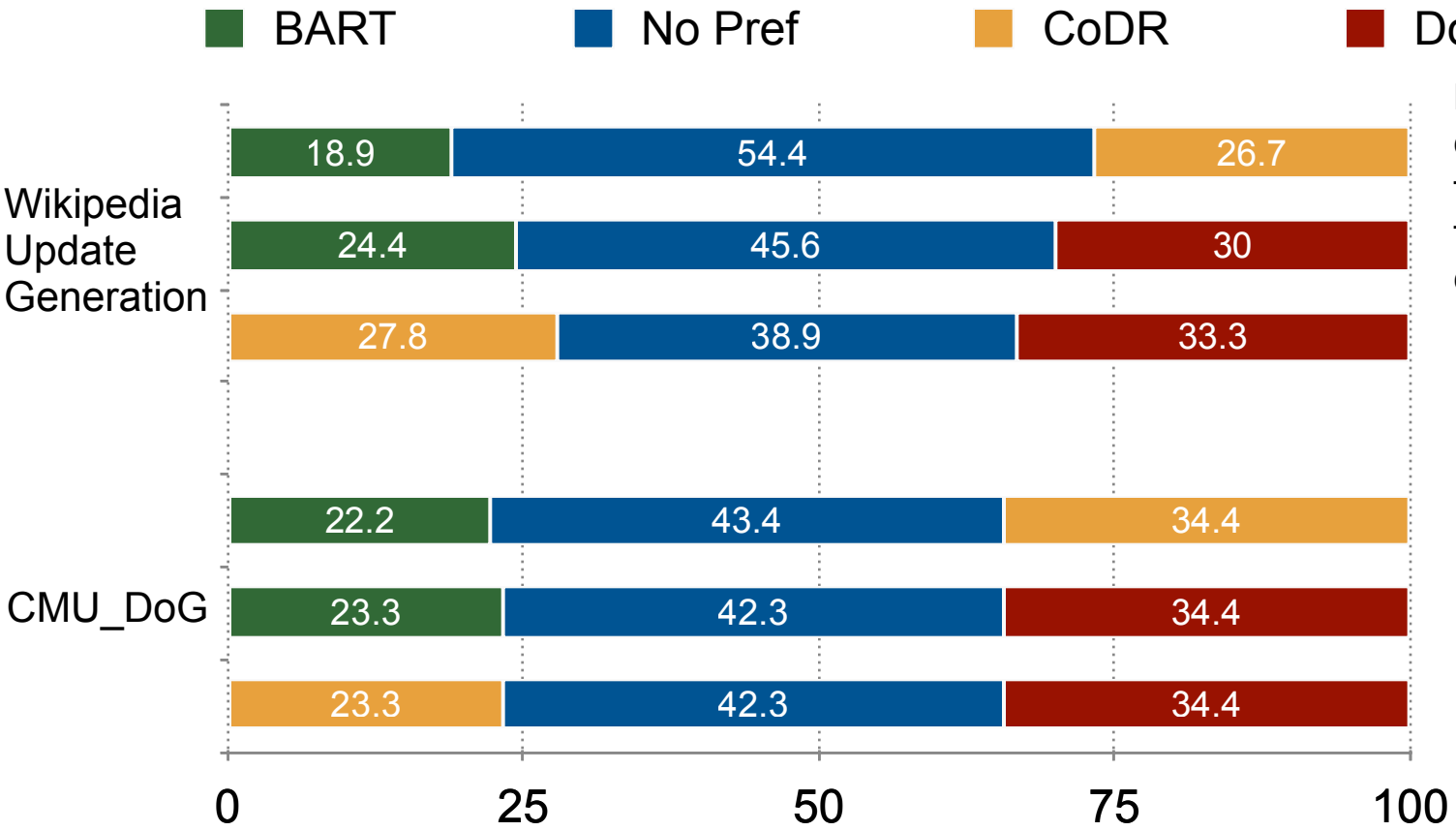

30
22.5
15
7.5
0
Low-Res (Zhao et al, ICLR 2020)
BART
CoDR
DoHA
• 19.7 BLEU-4 improvement over Low-Res
• 24% BLEU-4 improvement over BART
1.2
16.91
20.68
20.9
12.59
14.47
14.55
19.3
20.41
20.44
BLEU-4
METEOR
Rouge-L

Relevance Human Evaluation



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

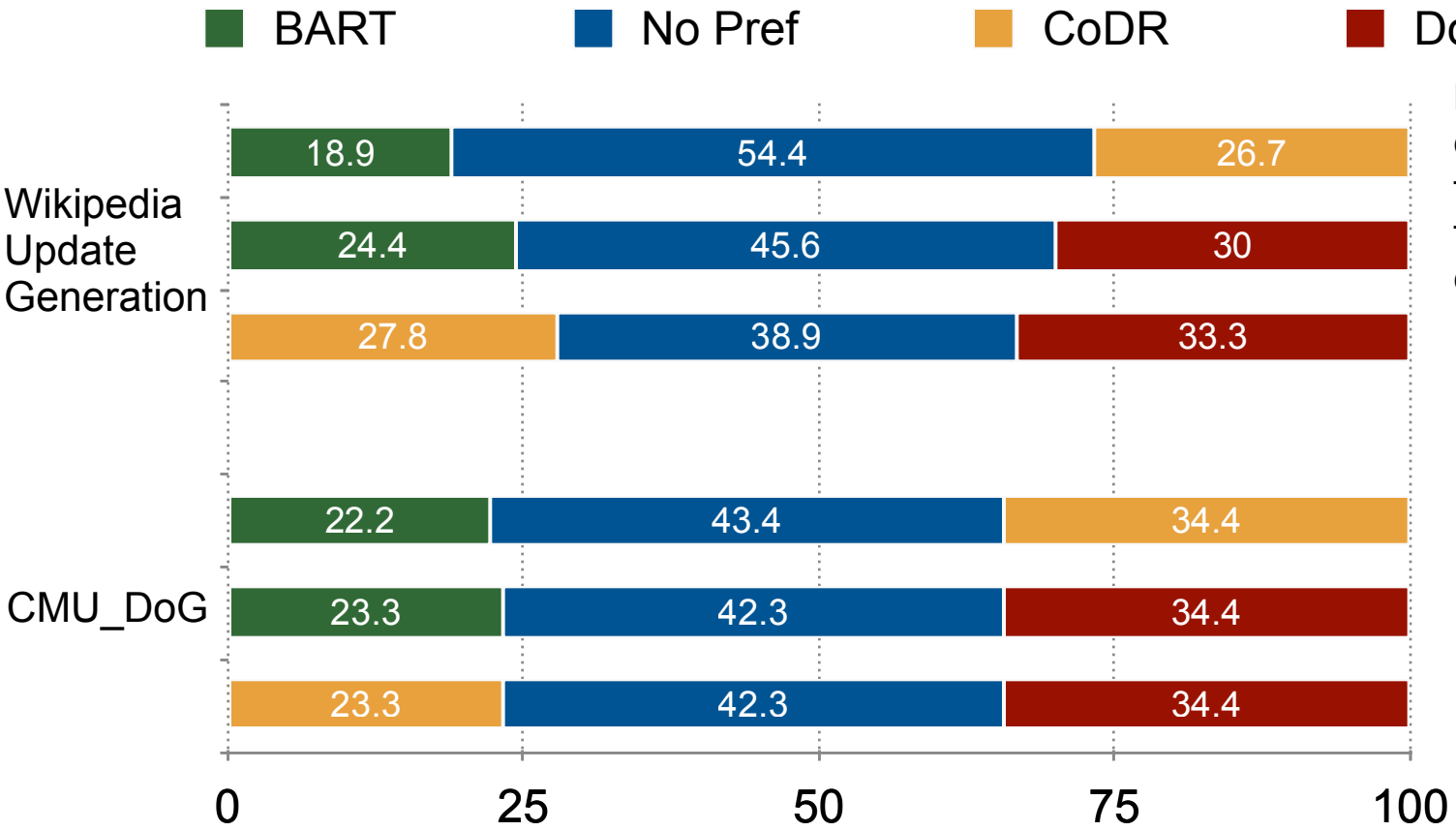
Relevance Human Evaluation



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
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
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
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
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

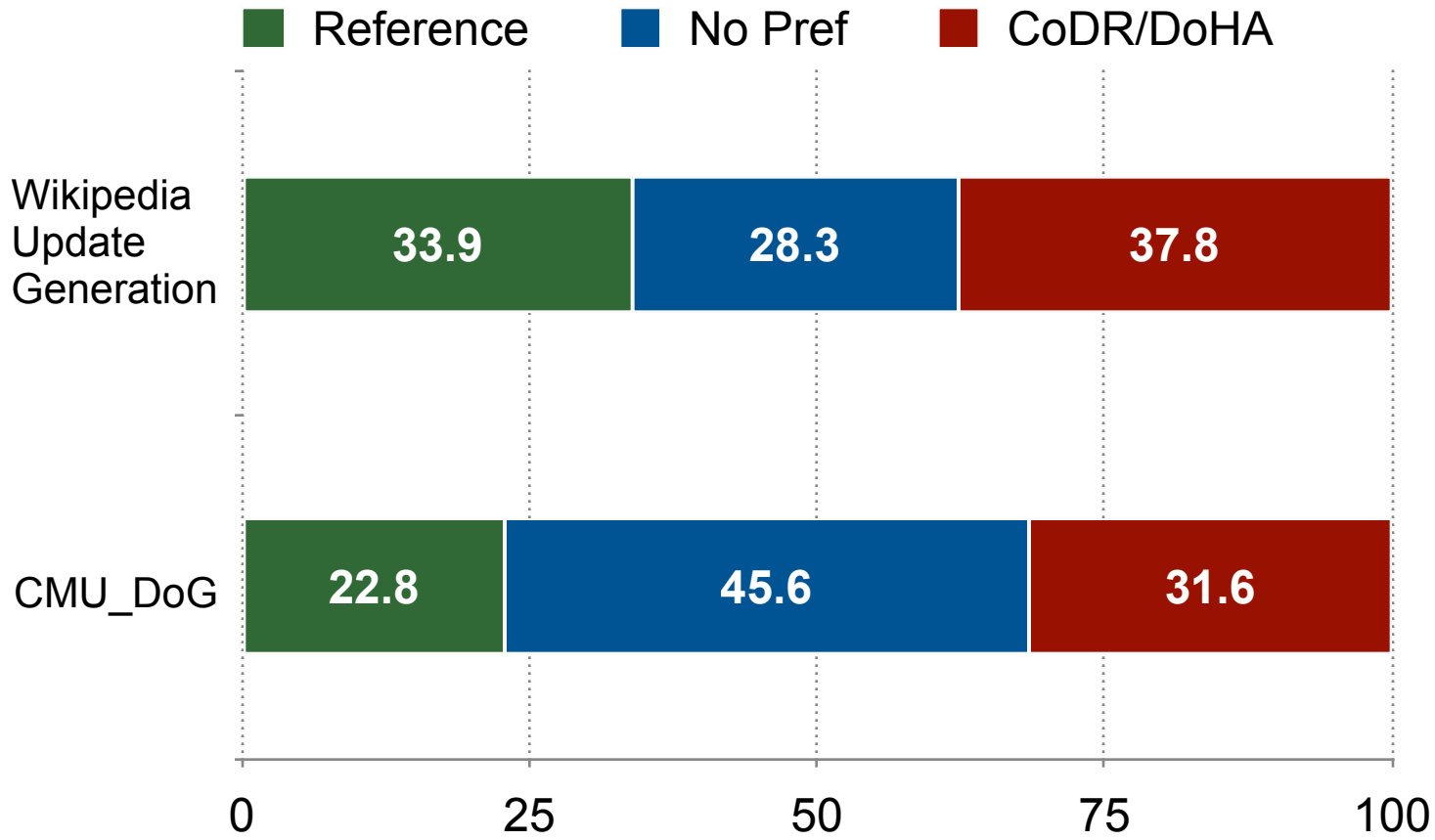
Error Class	%
Reference and generation are grounded	35
Reference is grounded but generation is not	3
Reference is not grounded but generation is	5
Reference and generation are not grounded	43

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

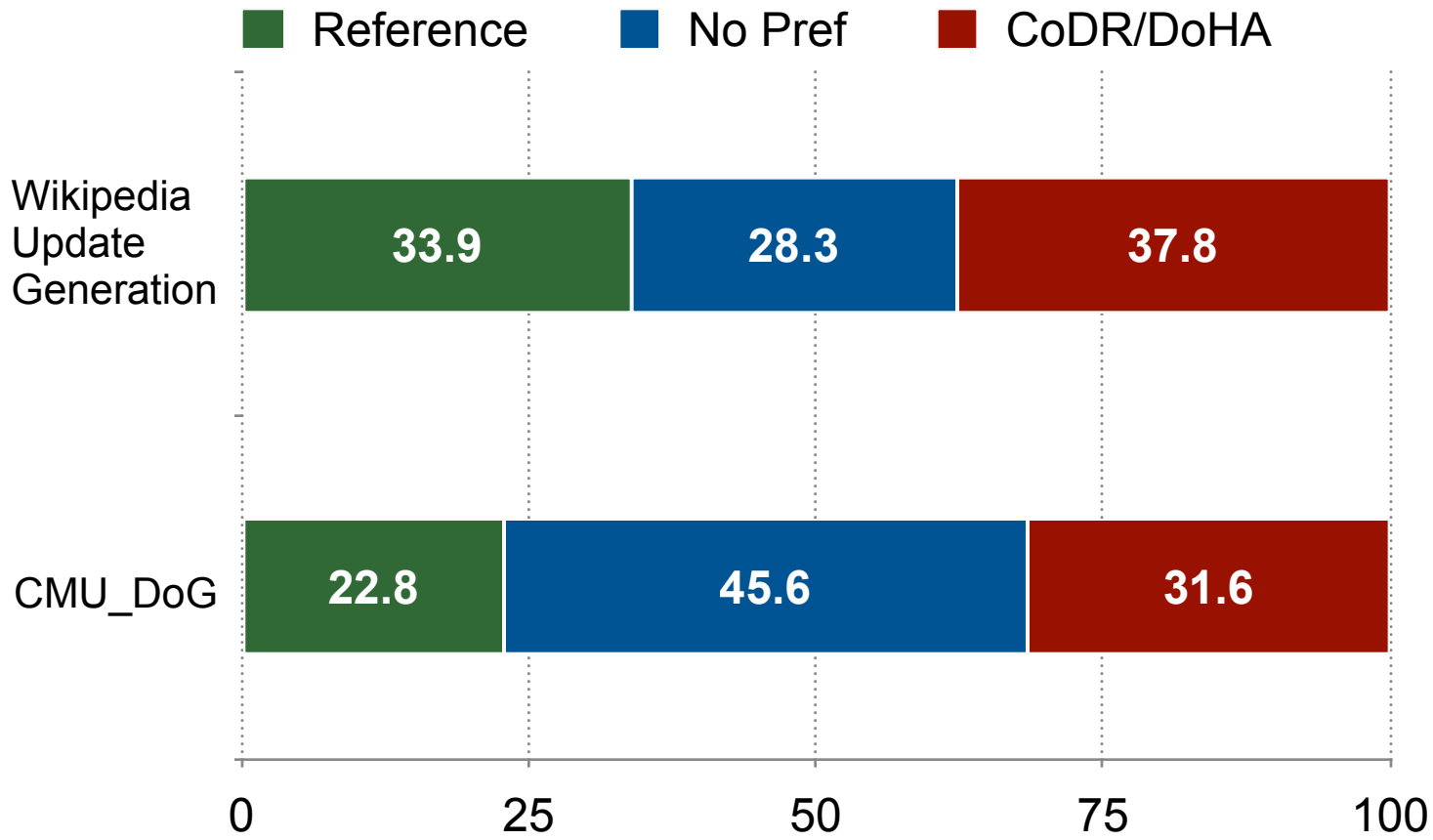
1. References are not grounded if they are follow up ***questions, opinions or experiences.***
2. ***Context dictates*** grounding
3. ***Conversational subtleties not captured*** by the automated metrics
4. ***90%*** times the generation is grounded if the reference is grounded and generation is not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

Ethical Considerations

Wiki Context	Document	Generation
Climate change includes both the global warming driven by human emissions … Though there have been previous periods of climatic change, since the mid-20th century the rate of human impact … impact have been unprecedented.	Rising sea levels? Well, it has not happened yet, but … I personally doubt it and … climate does not fit the narrative, it is called climate change, is a big money making hoax.	Though most scientists agree that a "global" climate change is occurring, climate change remains a contentious topic and many scientists, climate scientists, and even climate skeptics believe that it is a hoax.
	31,000 scientists reject global warming and say "no convincing evidence" that humans can or will cause global warming?	Some scientists dismiss the existence of climate change.
	Yes, you read that correctly, three million — *million* — years ago CO2 levels on Earth were the same as they are today,…Three million years ago, we humans were not driving cars …	The Scientific opinion on climate change is that planet-wide warming has occurred at least three million years ago.

Ethical Considerations

Wiki Context	Document	Generation
Climate change includes both the global warming drive by human emissions Though there have been previous periods of climatic change, since the mid-20th century the rate of human impact … impact have been unprecedented.	Rising sea levels? Well, it has not happened Yes, you read that correctly, three million — *million* — years ago CO2 levels on Earth were the same as they are today,…Three million years ago, we humans were not driving cars …	Though most scientists agree that a "global" climate change is occurring, mains a contentious scientists, climate en climate skeptics hoax. ismiss the existence of The Scientific opinion on climate change is that planet-wide warming has occurred at least three million years ago.

1. Model is sensitive to the document provided
2. Model does not fact check document provided
3. Model generates information faithful to the document

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

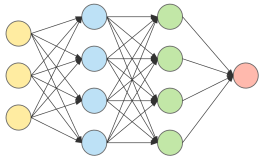
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

Common Themes

Overview

Style

ACL ’20
Storytelling ’19
ACL ’18

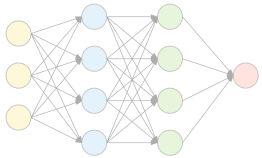
Content

NAACL ’21
NAACL ’19
EMNLP ’18

Structure

ACL ’20

Control Attributes



Modeling

COLING ’20

Ethics

NAACL ’21
chap. in L.L.C ’21
WiNLP ’19

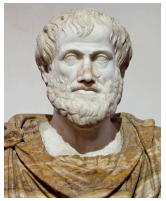
Common Themes

Ethics

- Prior work on understanding and mitigating bias (Hovy & Prabhumoye, 2021; Blodgett et al, 2020; Shah et al, 2020; Sun et al, 2019; Zhao et al, 2019; Tatman, 2017; Bolukbasi et al, 2016)

Ethics

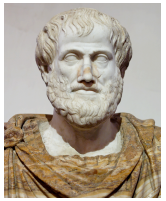
- Prior work on understanding and mitigating bias (Hovy & Prabhumoye, 2021; Blodgett et al, 2020; Shah et al, 2020; Sun et al, 2019; Zhao et al, 2019; Tatman, 2017; Bolukbasi et al, 2016)



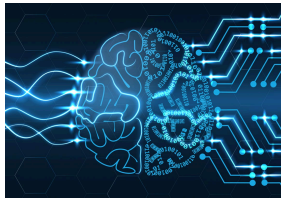
Large body of work on Ethics

Ethics

- Prior work on understanding and mitigating bias (Hovy & Prabhumoye, 2021; Blodgett et al, 2020; Shah et al, 2020; Sun et al, 2019; Zhao et al, 2019; Tatman, 2017; Bolukbasi et al, 2016)



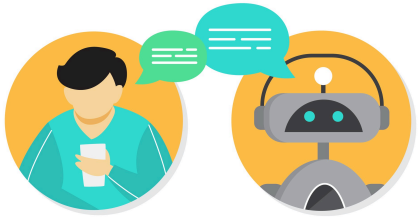
Large body of work on Ethics



How can we apply it to AI?

Ethics

- Deontological framework for NLP
  - Generalization principle and Respect for Autonomy
  - *Which tasks have important ethical implications?*
  - *What factors and methods are preferable in ethically solving this problem?*



Question-Answering

Machine Translation

Detecting objectionable content

Dialogue Systems

Generalization Principle

Generalization Principle

An action $\mathscr{A}$ taken for reasons $\mathscr{R}$ is ethical if and only if a world where all people perform $\mathscr{A}$ for reasons $\mathscr{R}$ is conceivable.

Generalization Principle

*An action $\mathscr{A}$ taken for reasons $\mathscr{R}$ is **un**ethical if and only if a world where all people perform $\mathscr{A}$ for reasons $\mathscr{R}$ logically contradicts $\mathscr{R}$*

Detecting objectionable content

Detecting objectionable content

- deploying flagging systems

Detecting objectionable content

$\mathscr{A}$ deploying flagging systems

$\mathscr{R}$ ↓ burden on humans ↓ # posts that need to be seen by human eyes

Detecting objectionable content

$\mathscr{A}$ deploying flagging systems

$\mathscr{R}$ ↓ burden on humans ↓ # posts that need to be seen by human eyes

"I like to imagine you as a girl but your sentence structure and rhetoric is so concise and to the point which points to the contrary (nothing against women, simply factual)."

Hate Speech Detection

API :: Perspective

Unlikely to be perceived as toxic (0.23)

Sentiment Analysis

python NLTK

Subjectivity
- neutral: 0.1
- **polar: 0.9**

Polarity
- **pos: 0.5**
- neg: 0.5

The text is **pos**.

[Breitfeller et al, EMNLP 2019]

Detecting objectionable content

$\mathscr{A}$ deploying flagging systems

$\mathscr{R}$ ↓ burden on humans ↓ # posts that need to be seen by human eyes

"I like to imagine you as a girl but your sentence structure and rhetoric is so concise and to the point which points to the contrary (nothing against women, simply factual)."

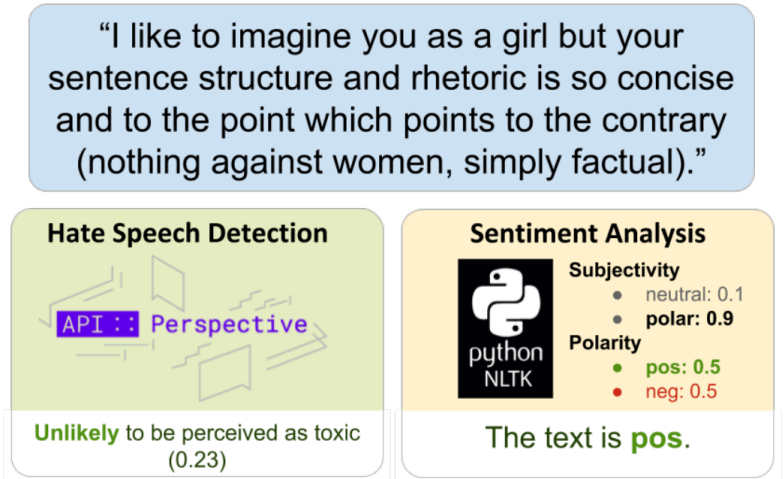
[Breitfeller et al, EMNLP 2019]

- surface level words ⟹ phrase the same meaning with different words

Detecting objectionable content

$\mathscr{A}$ deploying flagging systems

$\mathscr{R}$ ↓ burden on humans ↓ # posts that need to be seen by human eyes



[Breitfeller et al, EMNLP 2019]

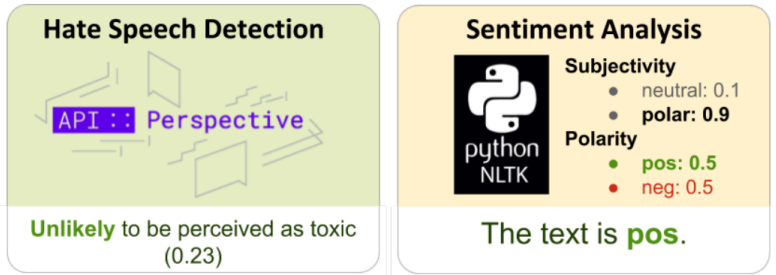
- surface level words $\Longrightarrow$ phrase the same meaning with different words
- flagging system will be unsuccessful

Detecting objectionable content

$\mathscr{A}$ deploying flagging systems

$\mathscr{R}$ ↓ burden on humans ↓ # posts that need to be seen by human eyes

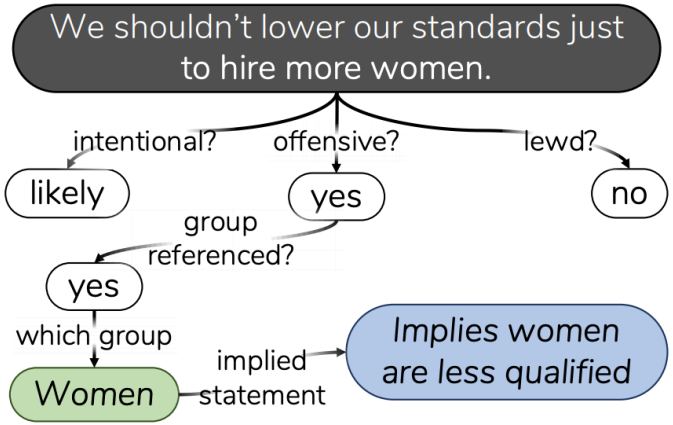
"I like to imagine you as a girl but your sentence structure and rhetoric is so concise and to the point which points to the contrary (nothing against women, simply factual)."



[Breitfeller et al, EMNLP 2019]

- surface level words ⟹ phrase the same meaning with different words
- flagging system will be unsuccessful
- logically contradicts the premise

Detecting objectionable content



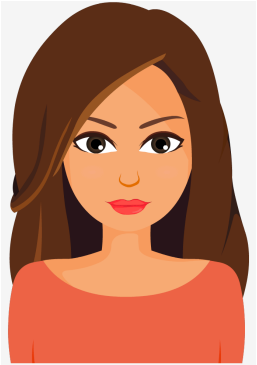
[Sap et al, ACL 2020]

- Underlying intent, offensiveness, and power differentials between different social groups.
- Controllable Text Generation
  - generate consequences and implications
- Does not lead to an arms race between objection content generation and detection

Respect for Autonomy

- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***

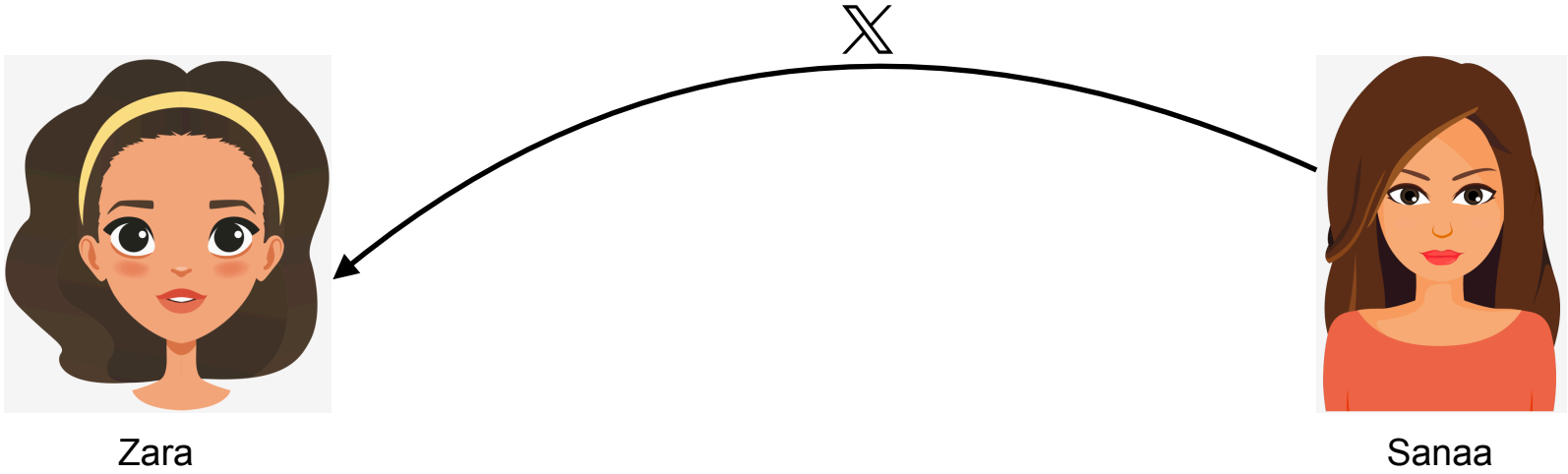
Zara



Sanaa

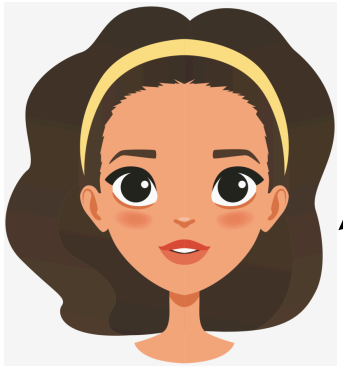
Respect for Autonomy

- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***

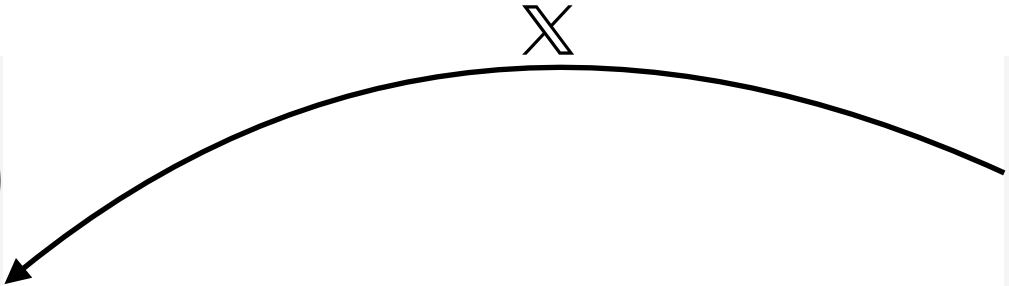


Respect for Autonomy

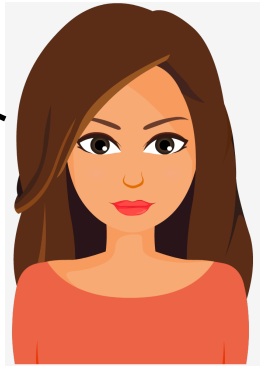
- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***



Zara



infringe on Zara's right to self-govern

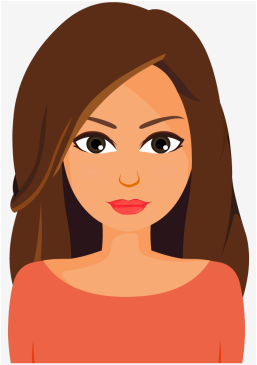


Sanaa

Respect for Autonomy

- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***

Zara



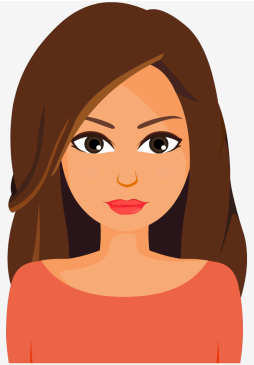
Sanaa

Respect for Autonomy

- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***

Zara

Zara must be sufficiently informed about $\mathbb{X}$



Sanaa

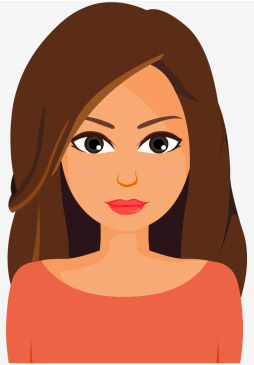
Respect for Autonomy

- Addresses the right of a person to make decisions which directly pertain to themselves.
- ***Informed consent***

Zara

Zara must be sufficiently informed about 𝕏

Zara *herself* makes the decision to allow Sanaa to do 𝕏

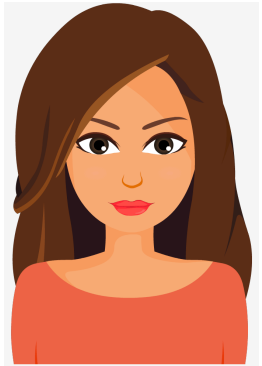


Sanaa

Translation

Translator

Zara



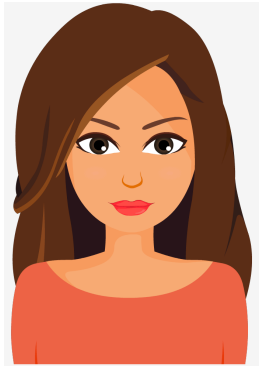
Sanaa

Translation

Zara consents to Sanaa serving as an *ad hoc* representative for what she would like to say.

Zara

Translator



Sanaa

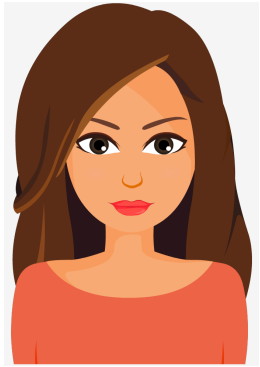
Translation

Zara consents to Sanaa serving as an *ad hoc* representative for what she would like to say.

There might be a formal contract of how Sanaa is to act

Zara

Translator



Sanaa

Translation

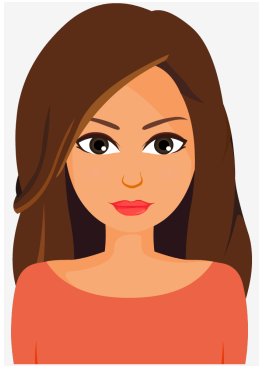
Zara consents to Sanaa serving as an *ad hoc* representative for what she would like to say.

Translator

Zara

There might be a formal contract of how Sanaa is to act

Zara relies on Sanaa's paralinguistic conduct



Sanaa

Machine Translation

Zara

Translator

Machine Translation

Machine Translation

MT system is speaking for Zara

Zara

Translator

Machine Translation

Machine Translation

MT system is speaking for Zara

Zara must be ***informed*** of ambiguities so that she can ***consent*** to the message which the system ultimately conveys.

Zara

Translator

Machine Translation

Machine Translation

MT system is speaking for Zara

Zara must be ***informed*** of ambiguities so that she can ***consent*** to the message which the system ultimately conveys.

Zara must also be ***informed*** of the failure cases in the MT system.

Zara

Translator

Machine Translation

Machine Translation

Zara

Ms. Hashimoto …

Translator

English to Japanese Machine Translation

-san? or
-sensei?…

Machine Translation

Zara must be notified that such an ambiguity needs to be resolved because there is a risk of offending the Japanese speaker.

Zara

Ms. Hashimoto …

Translator

English to Japanese Machine Translation

-san? or
-sensei?…

Machine Translation

My *aunt* is coming home tomorrow.

Zara

Translator

English to Hindi Machine Translation

Is it maternal or paternal aunt? They have different words in Hindi.

Aadil

Machine Translation

MT system can ask a follow up question to Zara.

Zara

My *aunt* is coming home tomorrow.

Translator

English to Hindi Machine Translation

Is it maternal or paternal aunt? They have different words in Hindi.

Aadil

Machine Translation

Zara

My *aunt* is coming home tomorrow.

Translator

English to Hindi Machine Translation

I am unable to translate the sentence in its current form. Can you please rephrase it?

Aadil

Ethical impact of Controllable Text Generation

Machine Translation: understand social context, control formality, politeness, author attributes, voice

Ethical impact of Controllable Text Generation

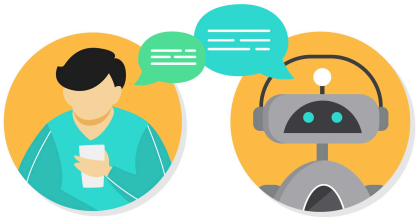
Machine Translation: understand social context, control formality, politeness, author attributes, voice

Detecting objectionable content: generate consequences and implications

Ethical impact of Controllable Text Generation

Machine Translation: understand social context, control formality, politeness, author attributes, voice

Detecting objectionable content: generate consequences and implications

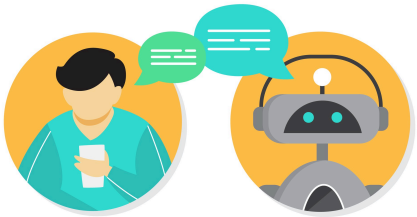


Question-Answering: transparency, dynamic graph generation for answers

Ethical impact of Controllable Text Generation

Machine Translation: understand social context, control formality, politeness, author attributes, voice

Detecting objectionable content: generate consequences and implications



Question-Answering: transparency, dynamic graph generation for answers

Dialogue Systems: control topics, style, content, persona

The key to human-like generation is Style, Content and Structure

Style

1. Back-Translation
2. Tag and Generate

Models

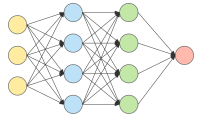
ACL '20
Storytelling '19
ACL '18

1. Politeness
2. Gender
3. Political Slant
4. Persona for Visual Stories

Tasks

1. Schema
2. Overview

COLING '20



Modeling

Content

1. CoDR
2. DoHA

Models

NAACL '21
NAACL '19
EMNLP '18

1. Wikipedia Update Generation
2. Document Grounded Dialogue

Tasks

Structure

1. Topological Sort for Sentence Ordering

ACL '20

Models

1. Deontological ethics for NLP
2. Four Case Studies

NAACL '21
chap. in L.L.C '21
WiNLP '19

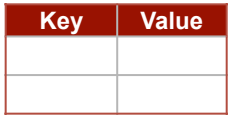
Ethics

Future Directions

- multi-lingual tasks

document

or



table

होनोलूलू में जन्में ओबामा किन्याई मूल के अश्वेत पिता व अमरीकी मूल की माता के संतान — Hindi

Родился в городе Гонолулу, штат Гавайи. — Russian

- evaluation

human evaluation is expensive

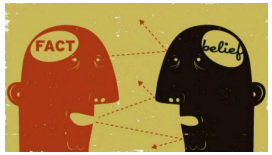
reproducible

task specific

automated metric

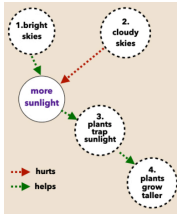
Future Directions

- expand control attributes



Bias

Persuasion



Commonsense reasoning

under review '21

- control multi-attributes

Style

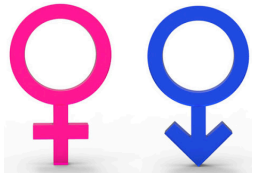
\+

Content

in progress

Age

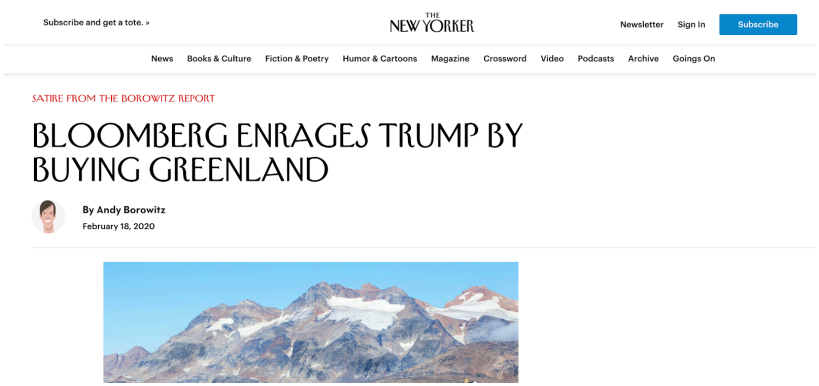
\+



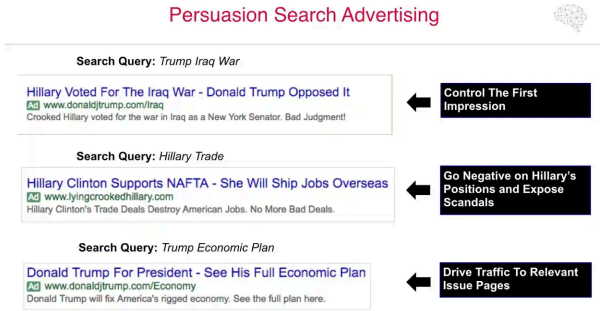
Gender

Impact on society

Propaganda: using stylized lexicons to make it appealing



Fake news: control content, personalize



Persuasion: control sentiment, strategy

Hate Speech: control target groups

Impact on society

"One size fits all" **in communication!**

Impact on society

Personalization: control target audience

Inclusion: accessible to all groups, control content

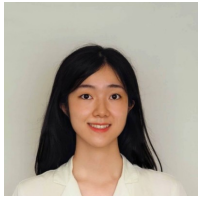
Anonymity: protect author attributes

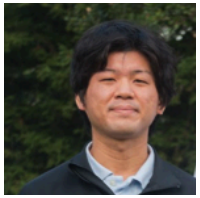
Balanced datasets: control demographic attribute

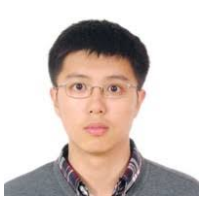
Thank You

Microsoft Research

Università Bocconi MILANO




salesforce research



facebook artificial intelligence research

Thank You

Jason Weston

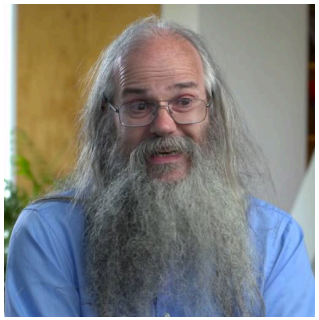
NYU

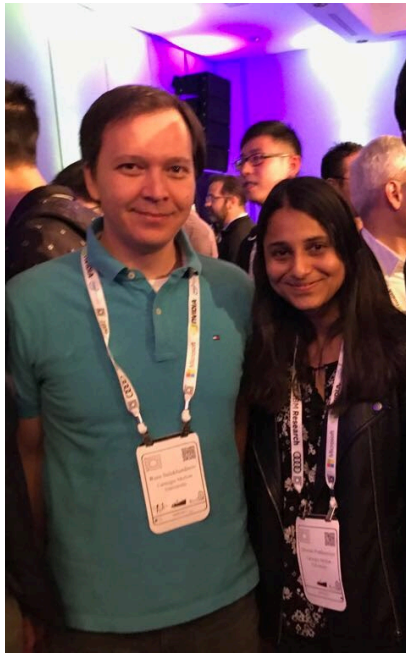
facebook artificial intelligence research

Thank You

Yulia Tsvetkov

Thank You





Alan W Black

Ruslan Salakhutdinov

The key to human-like generation is Style, Content and Structure

Style

1. Back-Translation
2. Tag and Generate

Models

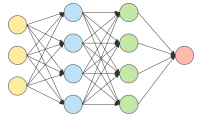
ACL '20
Storytelling '19
ACL '18

1. Politeness
2. Gender
3. Political Slant
4. Persona for Visual Stories

Tasks

1. Schema
2. Overview

COLING '20



Modeling

Content

1. CoDR
2. DoHA

Models

NAACL '21
NAACL '19
EMNLP '18

1. Wikipedia Update Generation
2. Document Grounded Dialogue

Tasks

Structure

1. Topological Sort for Sentence Ordering

ACL '20

Models

1. Deontological ethics for NLP
2. Four Case Studies

NAACL '21
chap. in L.L.C '21
WiNLP '19

Ethics