

Controllable Text Generation

Should machines reflect the way humans interact in society?

Shrimai Prabhumoye



Carnegie Mellon University

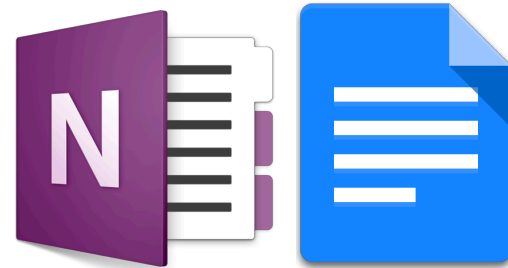
Language Technologies Institute

We constantly author text!

Style, Content and Structure



Social Media



Personal/Professional documents



Blogs



News/Wiki articles

Style

- Expressed in the choice of *words or phrases* as well as *syntactic structures* used to convey information
- I define style as a group of natural language sentences that belong to a particular class or label.

“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

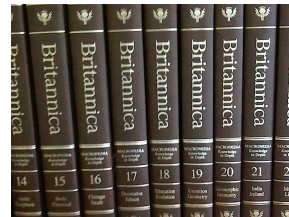
- Human communication by definition is a process by which individuals *exchange information* and influence one another through a common system of symbols and signs.
- Leverage information in the unstructured form



Wikipedia



Gutenberg



Encyclopedia



Internet

Content

Alice: The Notebook is hands-down one of my favorite movies EVER! Have you ever seen The Notebook?

Bob: No I have never seen this movie. I am going to try it out now

Alice: It was a heartwarming story of young love. The main characters are played by [Ryan Gosling and Rachel McAdams](#).

Bob: Ok this sounds nice. I think Ryan is a good actor

Alice: For all the praise it received, I was surprised to see that it only got a [5.7/10 on Rotten Tomatoes](#).

Bob: That is interesting. They never get the rating correct.

Alice: [The story goes back and forth between present day and the past. Older Ryan is played by James Garner and older Rachel is played by Gena Rowlands](#). Yeah, Rotten Tomatoes never gets the right ratings. I always like to see the ratings but if I want to see a movie, I will watch it even if it has a bad rating.

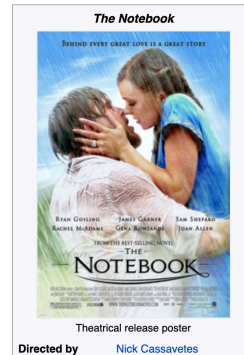
The Notebook

From Wikipedia, the free encyclopedia

For other uses, see *Notebook* (disambiguation).

The Notebook is a 2004 romantic drama film directed by Nick Cassavetes and written by Jeremy Leven from Jan Sardi's adaptation of the 1996 novel by Nicholas Sparks. The film stars Ryan Gosling and Rachel McAdams as a young couple who fall in love in the 1940s. Their story is narrated from the present day by an elderly man (played by James Garner), telling the tale to a fellow nursing home resident (played by Gena Rowlands, who is Cassavetes's mother).

The Notebook received generally mixed reviews, but performed well at the box office and received a number of award nominations, winning eight Teen Choice Awards, a Satellite Award, and an MTV Movie Award. The film became a sleeper hit^{[3][4]} and has gained a cult following.^{[5][6]} On November 11, 2012, ABC Family premiered an extended version with deleted scenes added back into the original storyline.^[7]



Reception [edit]

Box office [edit]

The film premiered June 25, 2004, in the United States and Canada and grossed \$13,464,745 in 2,303 theaters its opening weekend, ranking number 4 at the box office.^[33] The film grossed a total of \$115,603,229 worldwide, \$81,001,787 in Canada and the United States and \$34,601,442 in other countries.^[2] It is the 15th highest-grossing romantic drama film of all time.^[34]

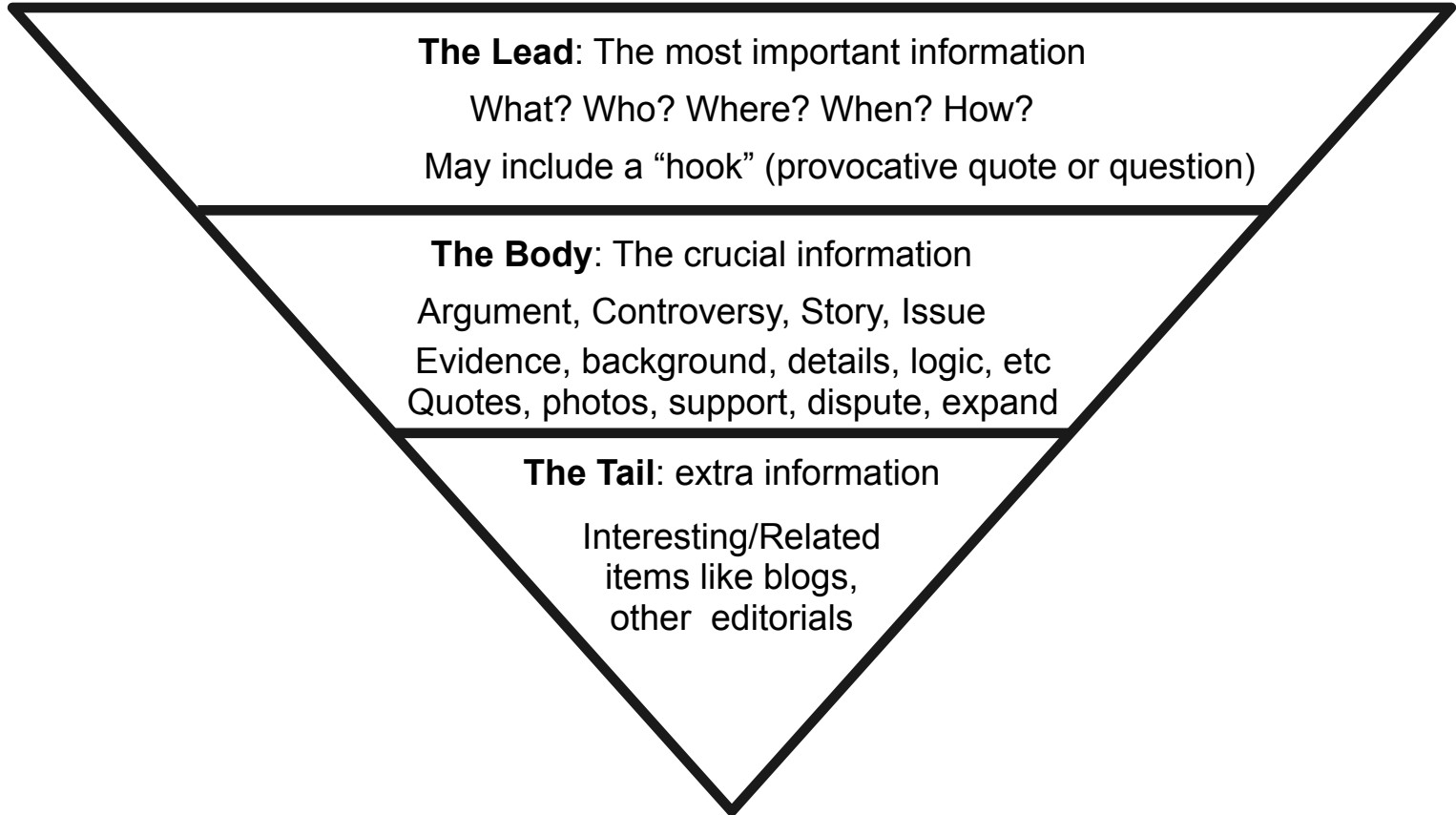
Critical reception [edit]

The Notebook received a mixed reaction from film critics. The 178 reviews on review aggregator *Rotten Tomatoes* show that 53% of critics gave the film a positive review, with an average rating of 5.64/10 and the website's consensus stating "It's hard not to admire its unabashed sentimentality, but *The Notebook* is too clumsily manipulative to rise above its melodramatic clichés."^[35] At *Metacritic*, which assigns an average rating out of 100 to reviews from mainstream critics, the film currently holds an average score of 53, based on 34 reviews, which indicates "mixed or average reviews."^[36]



The performances of Ryan Gosling and Rachel McAdams, as well as their on-screen chemistry, were particularly praised by most film critics

Structure

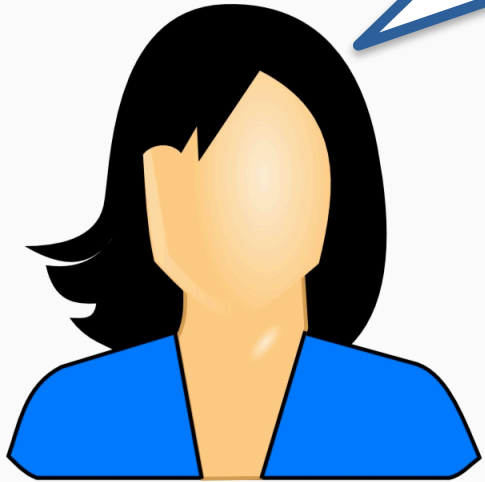


Application

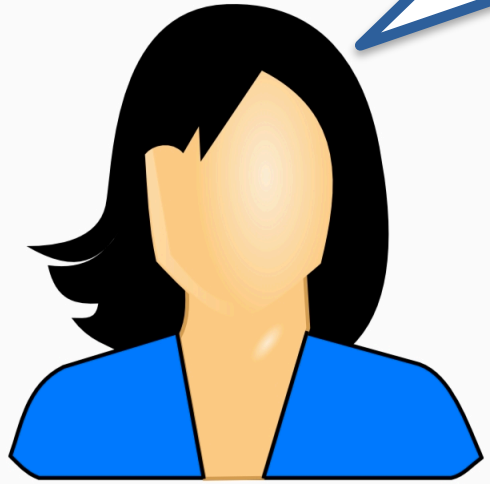


Application

I have a talk to
present today!



Application



I have a talk to present today!



Best of luck!

Application



I have a talk to
present today!




Best of luck!

Style

Application



I have a talk to
present today!



May the force be with
you!

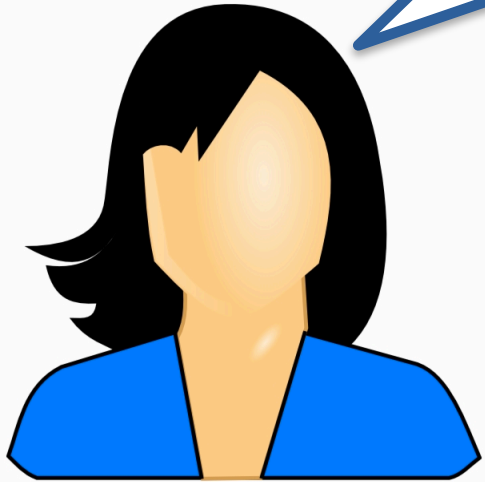
✓
Style

Application



Application

Hey, have you
watched Avengers
Endgame



Application



Hey, have you
watched Avengers
Endgame



Yeah, I loved Bradley
Cooper in it!

Application



Hey, have you
watched Avengers
Endgame



Yeah, I loved Bradley
Cooper in it!


Content

Application



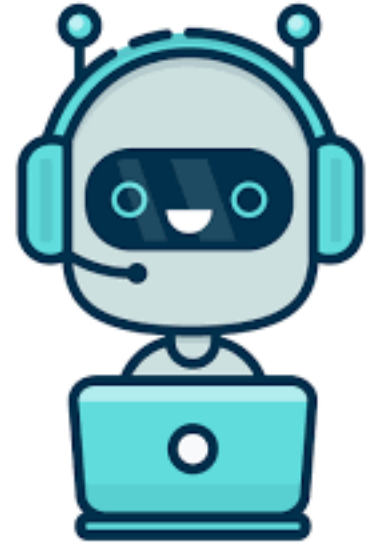
Hey, have you
watched Avengers
Endgame



Yeah, I loved Robert
Downey in it!

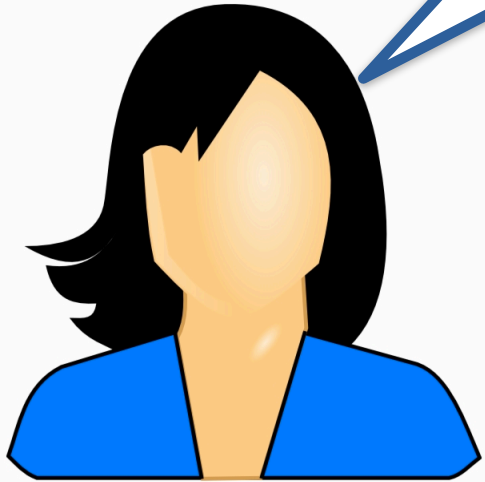
✓
Content

Application



Application

Hey, have you
watched Avengers
Endgame



Application



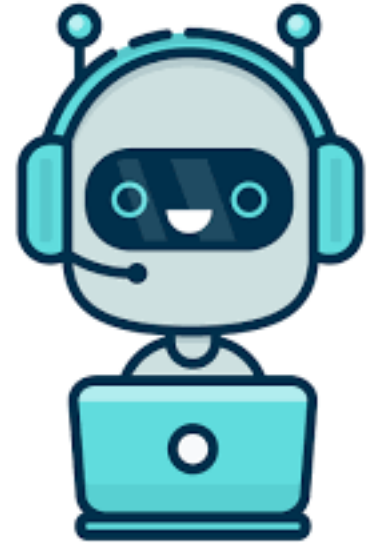
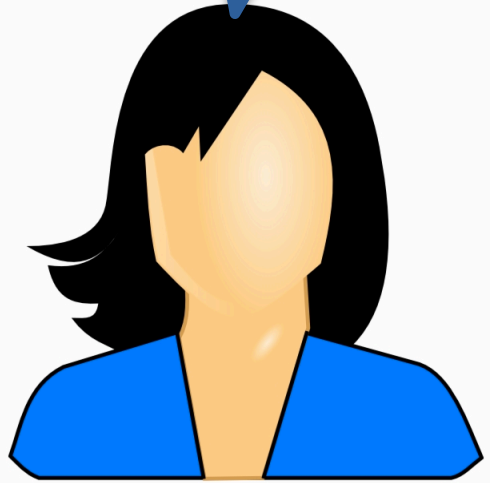
Hey, have you
watched Avengers
Endgame



Yeah, I loved the
movie!

Application

oh
great! Can you tell me
the story!



Application

oh
great! Can you tell me
the story!



Yeah, Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life. Thor decapitates Thanos. Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone. Five years later, AntMan escapes from the quantum realm. Ironman builds a time machine to save the world.



Application

oh
great! Can you tell me
the story!



Yeah, Ironman steals the
Infinity Stones back from Thanos
and uses them to disintegrate Thanos
and his army, at the cost of his life. Thor
decapitates Thanos. Hulk travels to New York
City in 2012 and convinces the Ancient One to
give him the Time Stone. Five years
later, AntMan escapes from the quantum
realm. Ironman builds a time
machine to save the world.

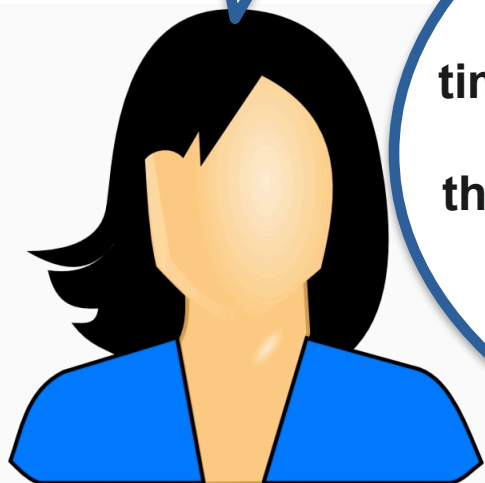


Structure



Application

oh
great! Can you tell me
the story!



Yeah, Thor ✓
decapitates Thanos. Five
years later, AntMan escapes from
the quantum realm. Ironman builds a
time machine to save the world. Hulk travels
to New York City in 2012 and convinces
the Ancient One to give him the Time Stone.
Ironman steals the Infinity Stones back
from Thanos and uses them to
disintegrate Thanos and his army,
at the cost of his life.

Structure

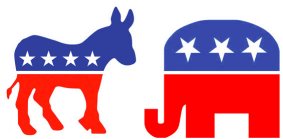


Other Applications

- ***Writing Assistance Tools***
 - recommend formal language
 - recommend structural changes
- Recommend ***polite emails***
- ***Story Generation***
 - plot, ending, sentiment, topic, persona
- ***Content Generation*** (websites, descriptions etc)

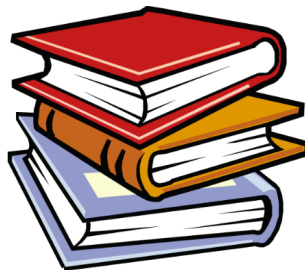
Overview

Style



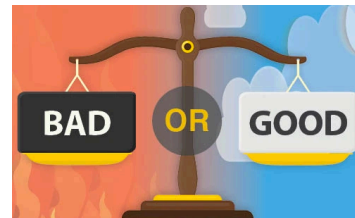
ACL '20
Storytelling '19
ACL '18

Content



NAACL '19
EMNLP '18

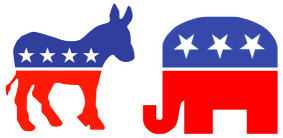
Ethical Considerations



WiNLP '19

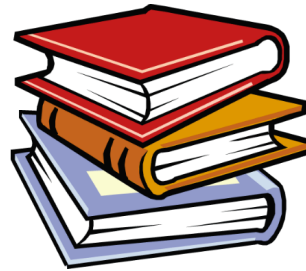
Overview

Style



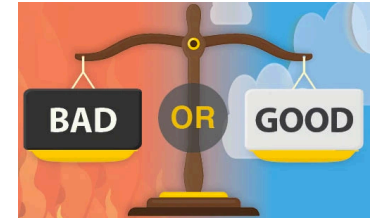
ACL '20
Storytelling '19
ACL '18

Content



NAACL '19
EMNLP '18

Ethical Considerations



WinLP '19

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.”

Applications

Anonymization: To preserve anonymity of users online, for personal security concerns (Jardine, 2016), or to reduce stereotype threat (Spencer, 1999).



Demographically-balanced training data for downstream applications.



Challenges

- No Parallel Data!

“The movie was very long.”

“I entered the theatre in the bloom of youth and emerged with a family of field mice living in my long, white mustache.”

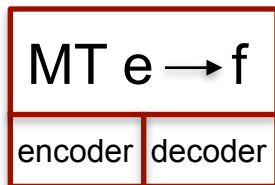
- Disentangle content from style
- Style is subtle

Our Solution

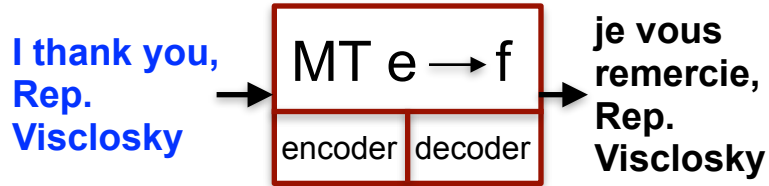
- ***Back-Translation***
 - Translating an English sentence to a pivot language and then back to English.
- Reduces stylistic properties
- Helps in grounding meaning
- Creates a representation independent of the generative model
- Representation is agnostic to the style task

Architecture

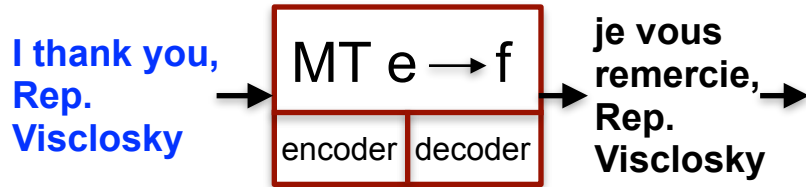
Architecture



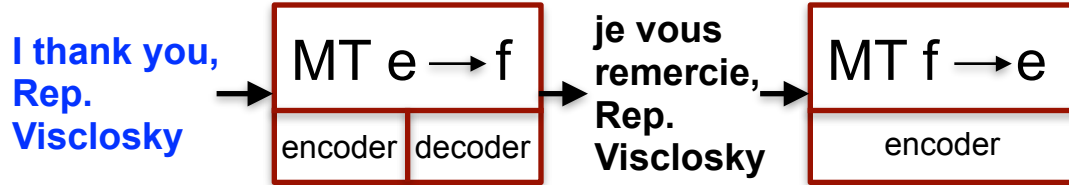
Architecture



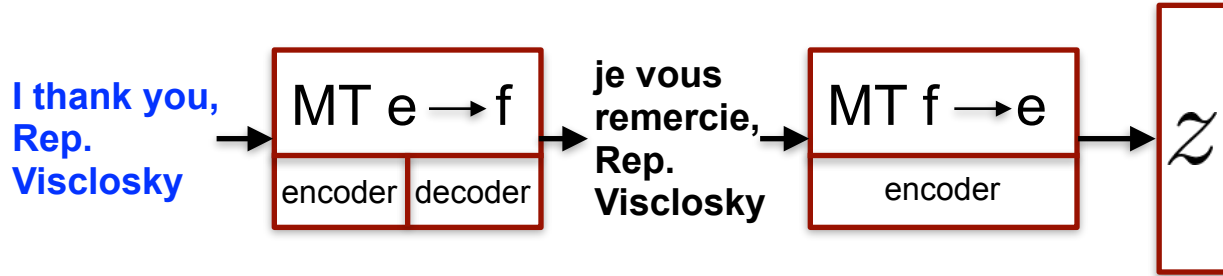
Architecture



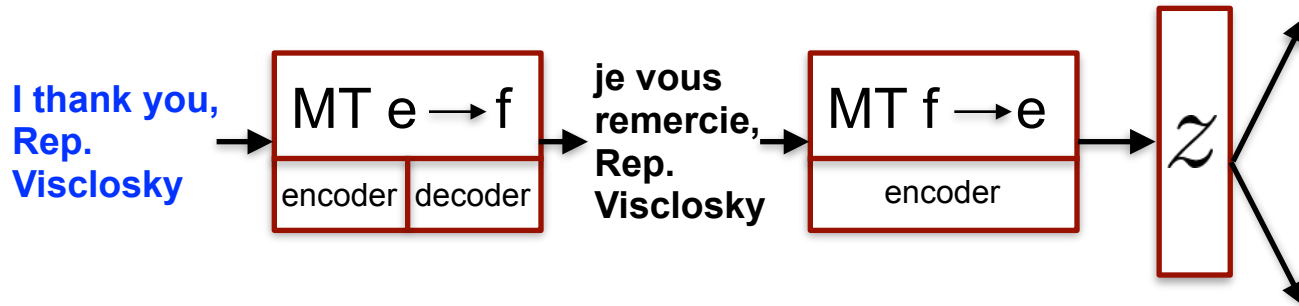
Architecture



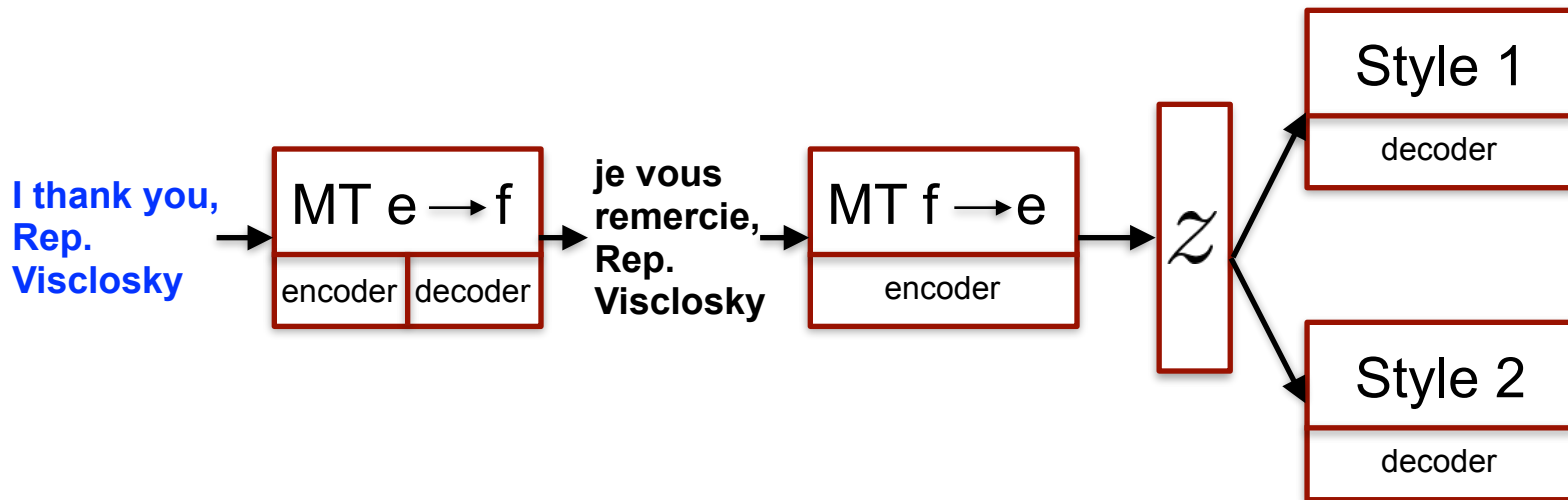
Architecture



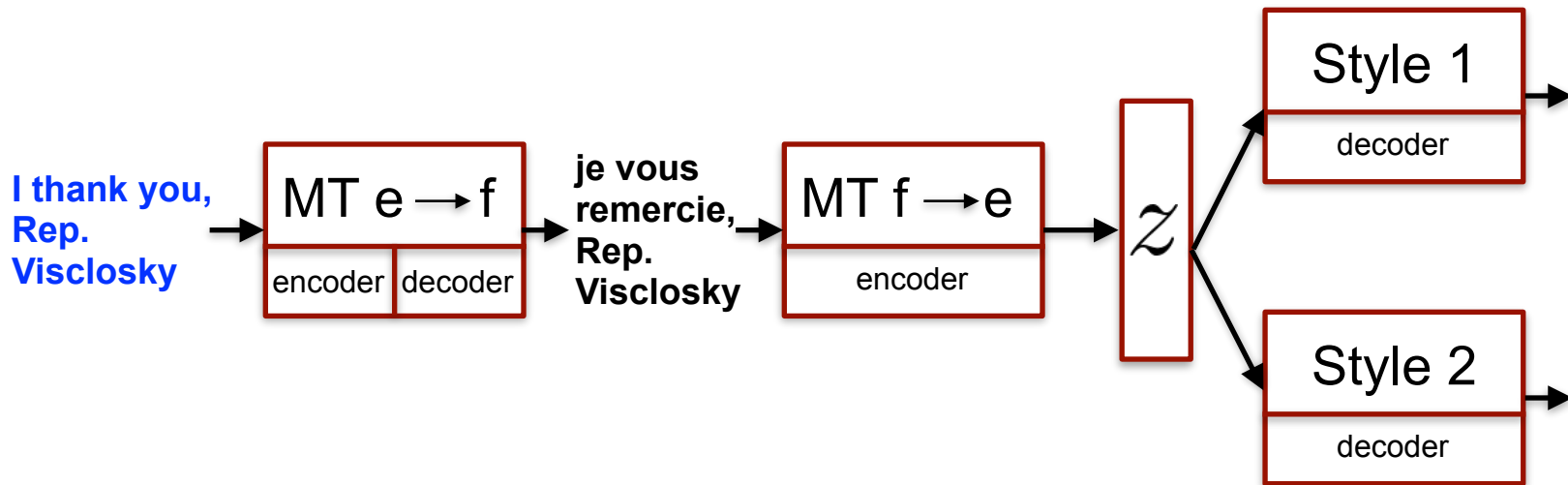
Architecture



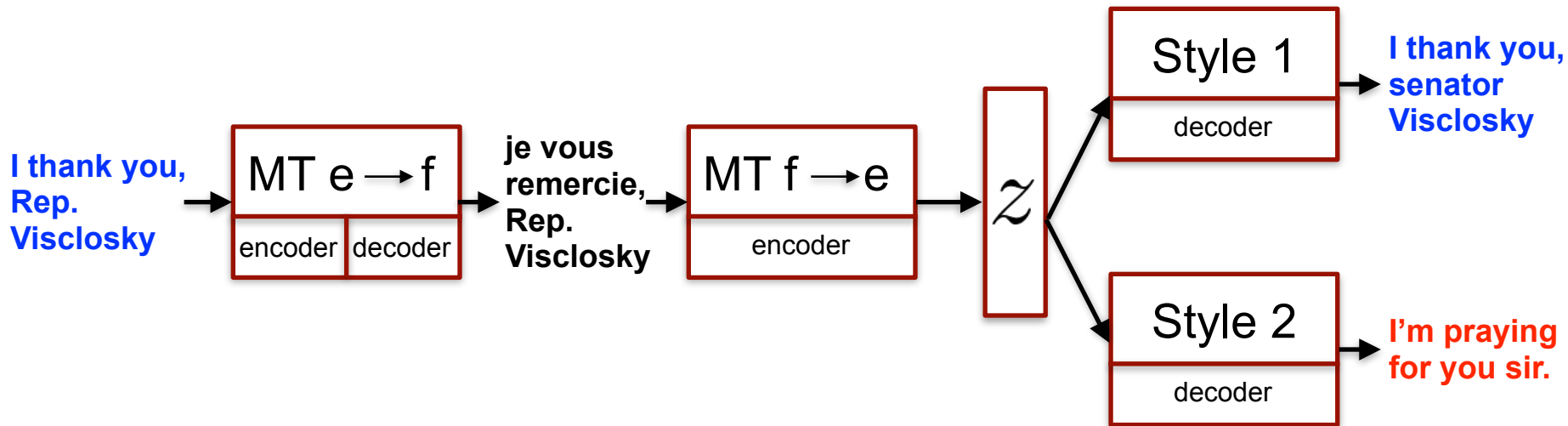
Architecture



Architecture

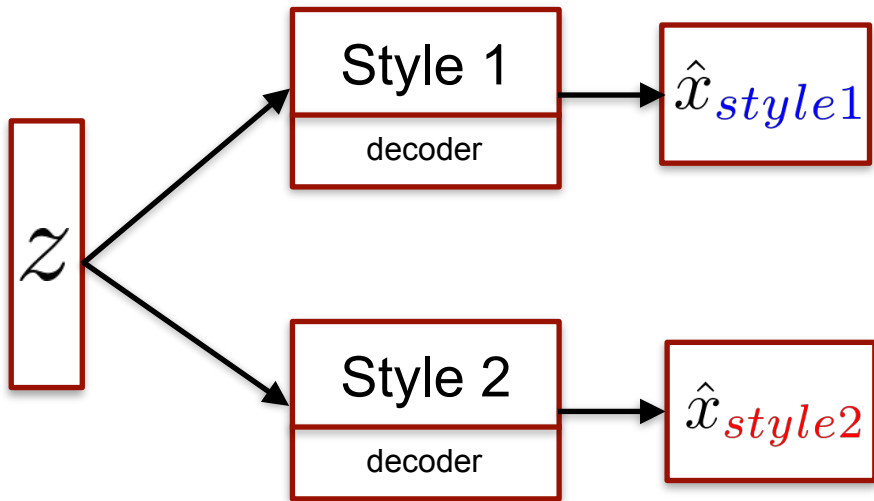


Architecture

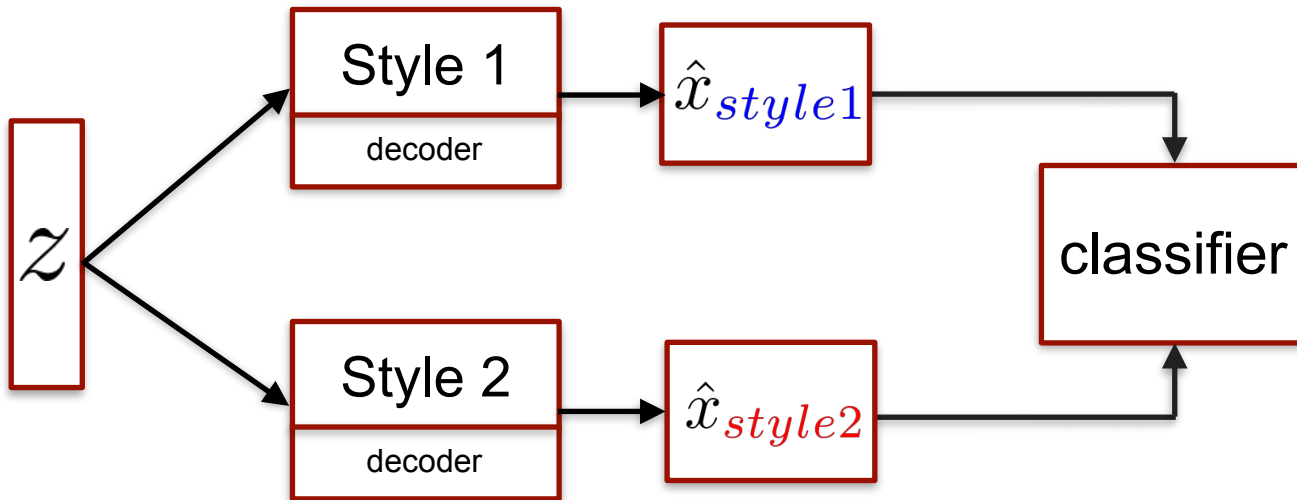


Train Pipeline

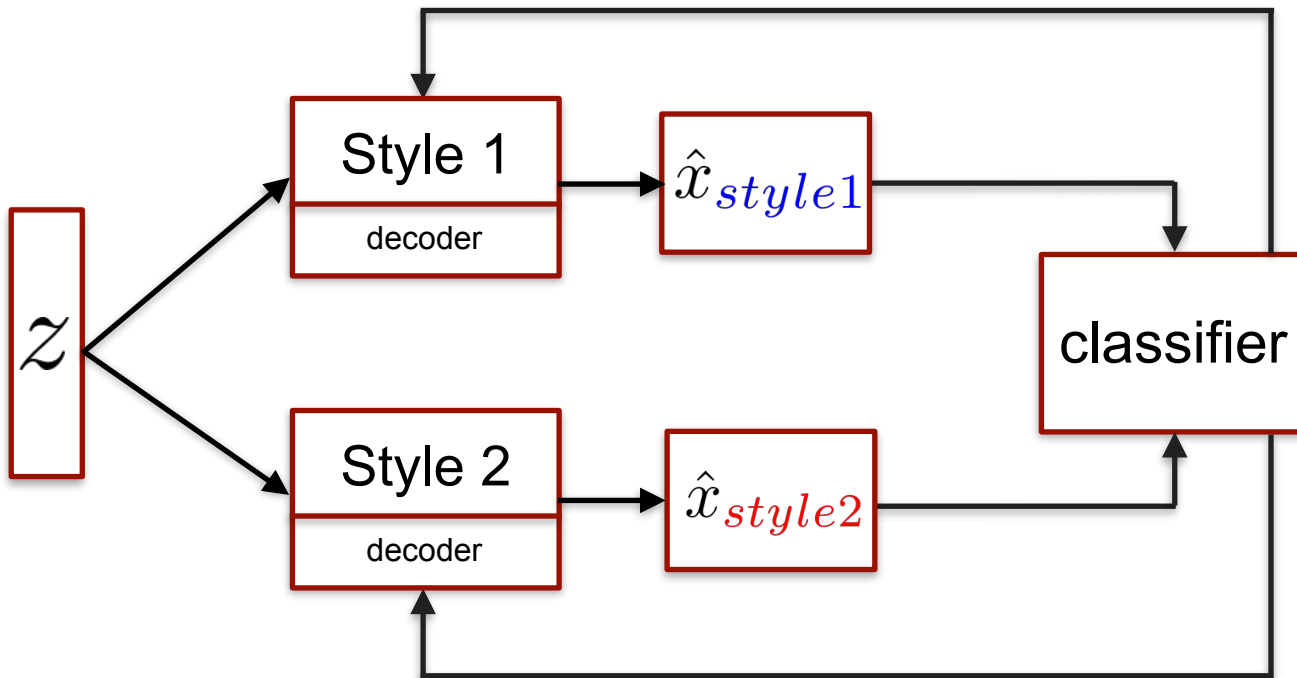
Train Pipeline



Train Pipeline



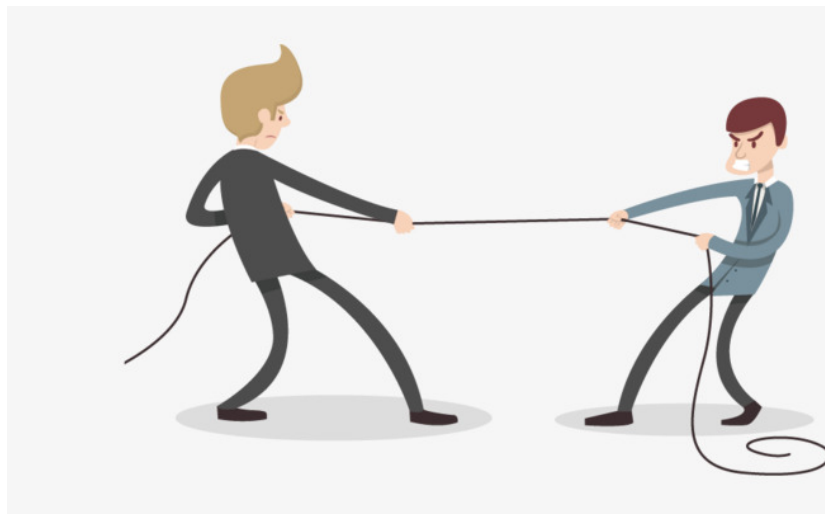
Train Pipeline



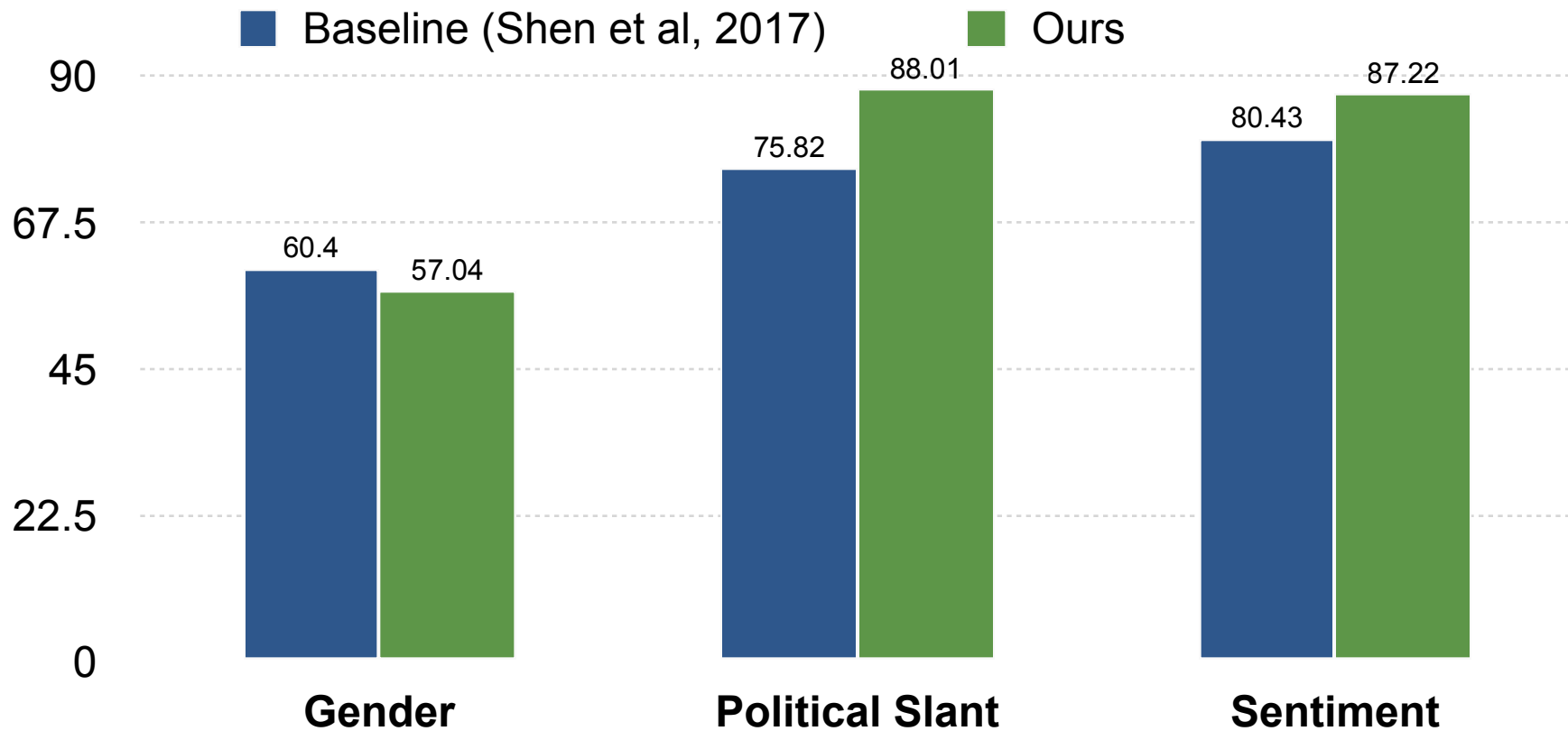
Experimental Settings

- Encoder-Decoders follow sequence-to-sequence framework (Sutskever et al., 2014; Bahdanau et al., 2015)

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



Style Transfer Accuracy



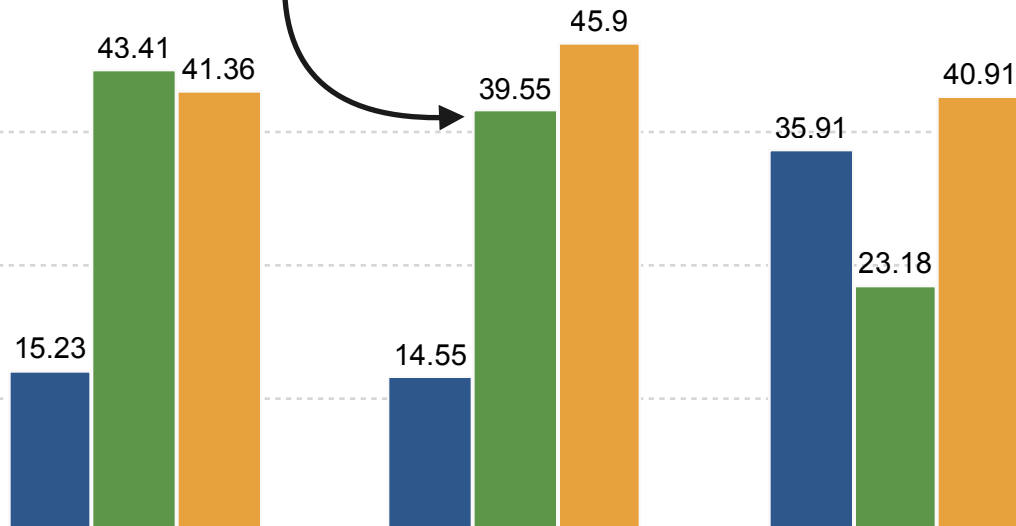
Preservation of Meaning

Which transferred sentence maintains the same semantic intent of the source sentence while changing the political position

- Baseline (Shen et al, 2017)
- Ours
- No Preference

Which transferred sentence maintains the same sentiment of the source sentence in the same semantic context (i.e. you can ignore if food items are changed)

50
37.5
25
12.5
0



Gender

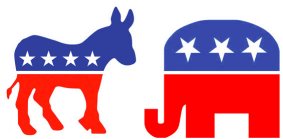
Political Slant

Sentiment

Which transferred sentence is semantically equivalent to the source sentence with an opposite sentiment

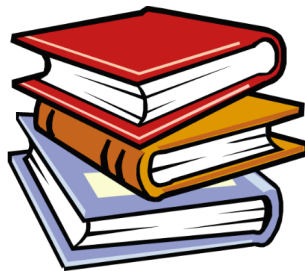
Overview

Style



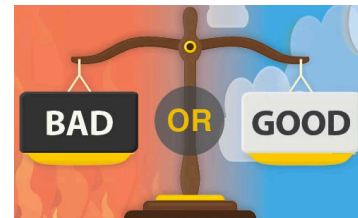
ACL '20 ✓
Storytelling '19
ACL '18

Content



NAACL '19
EMNLP '18

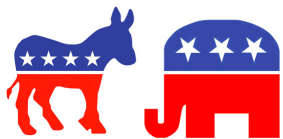
Ethical Considerations



WinNLP '19

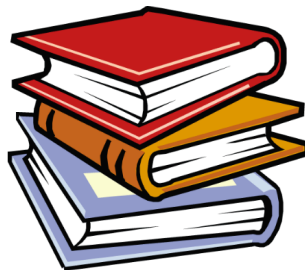
Overview

Style



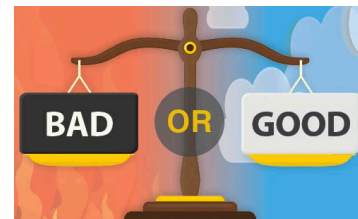
ACL '20 ✓
Storytelling '19
ACL '18

Content



NAACL '19
EMNLP '18

Ethical Considerations



WinNLP '19

Content Transfer

After **graduate form** Columbia University, Obama worked in Chicago.

After **graduating from** Columbia University, Obama worked in Chicago.

After graduating from **Carnegie Mellon University**, Obama worked in Chicago.

After graduating from **Columbia University**, Obama worked in Chicago.

- AI assistance deals with *form* (grammar, style, etc.)
- Our goal is to control for *content*

What is our task?

What is our task?



WIKIPEDIA
The Free Encyclopedia

Monkey selfie copyright dispute

From Wikipedia, the free encyclopedia



What is our task?

Background [edit]

On 4 July 2011 several publications including the *Daily Mail*,^{[8][10]} *The Telegraph*, and *The Guardian*^[11] picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: "Monkey steals camera to snap himself" (*The Telegraph*),^[12] "a camera on a tripod" triggered by the monkeys (*The Guardian*),^[13] and a camera started by a monkey "Fascinated by her reflection in the lens".^[10] The articles also contained Slater quotes such as "He must have taken hundreds of pictures by the time I got my camera back." 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.^[14]

spute



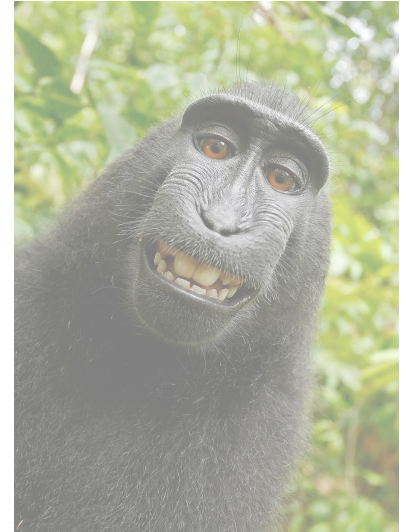
What is our task?



On 4 July 2011 several publications including the *Daily Mail*,^[8] ^[10] *The Telegraph*, and *The Guardian*^[11] picked up the story and published the pictures along with articles that quoted Slater as describing the photographs as self-portraits taken by the monkeys: "Monkey steals camera to snap himself" (*The Telegraph*),^[12] "a camera on a tripod" triggered by the monkeys (*The Guardian*),^[13] and a camera started by a monkey "Fascinated by her reflection in the lens".^[10] The articles also contained Slater quotes such as "He must have taken hundreds of pictures by the time I got my camera back." 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.^[14]

spute



What is our task?

On 4 July 2011 several publications including the *Daily Mail*,^[8] ^[10] *The Telegraph*, and *The Guardian*^[11] picked up the story and published the pictures along with articles that quoted Slater as

dispute

Ape-rture priority photographer plays down monkey reports

Chris Cheesman July 5, 2011

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

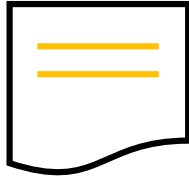
Wildlife photographer David Slater today played down newspaper reports that suggest a bunch of Indonesian monkeys grabbed his camera and began taking self-portraits.

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.^[14]



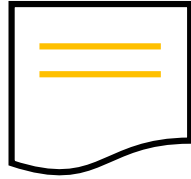
Primary Contribution

Primary Contribution



curated text (context)

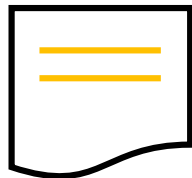
Primary Contribution



+

curated text (context)

Primary Contribution



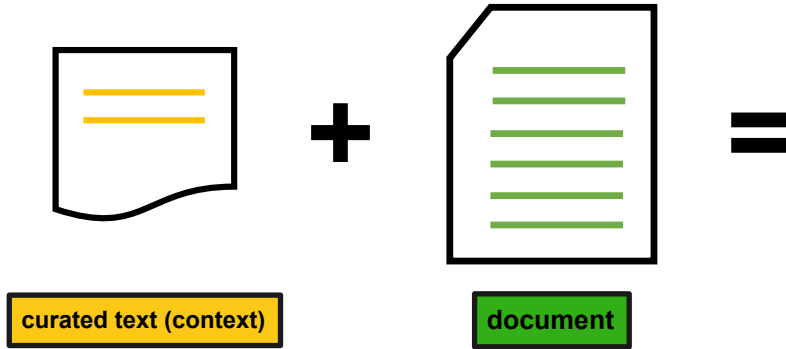
+



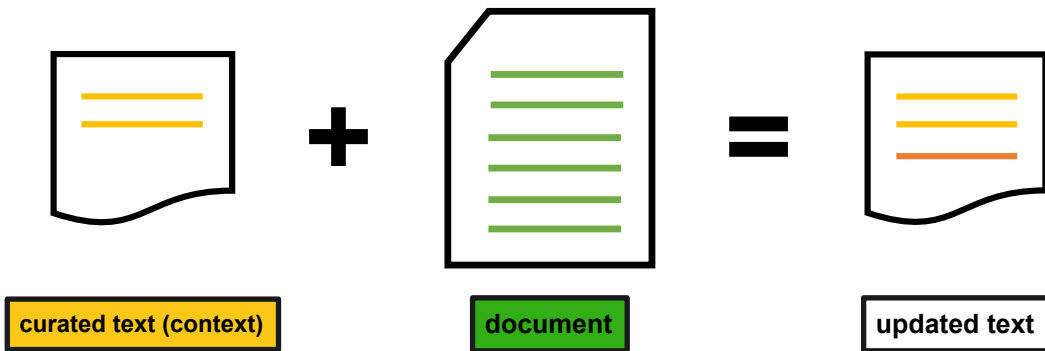
curated text (context)

document

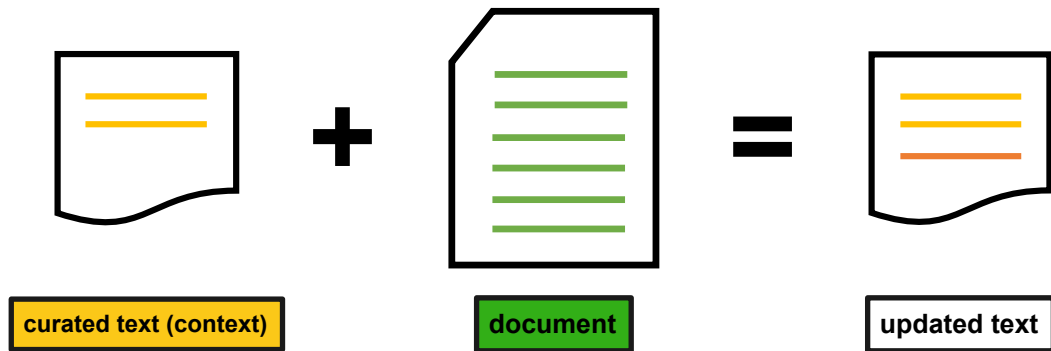
Primary Contribution



Primary Contribution



Primary Contribution

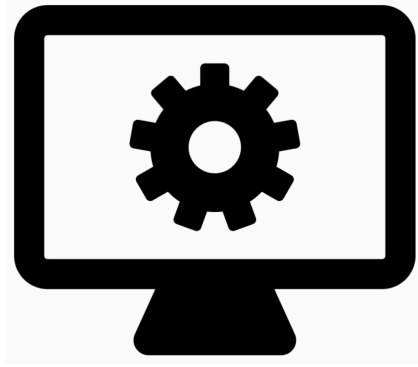


- design a task to perform content transfer from an unstructured source of information
- release dataset

Applications

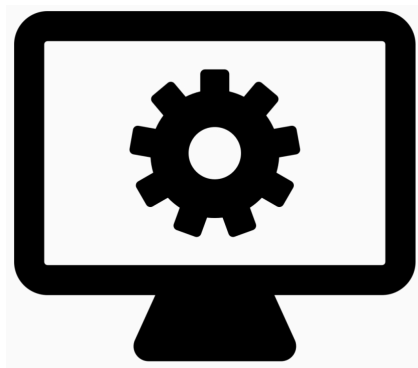
Applications

**Software
Documentation**



Applications

**Software
Documentation**

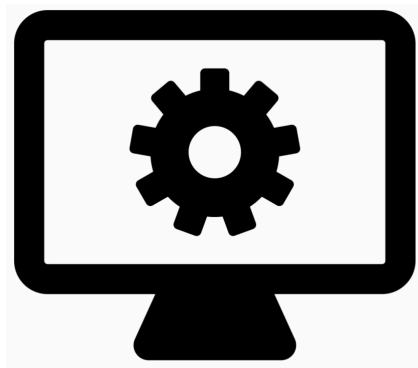


**Inbox
Summarization**



Applications

**Software
Documentation**



**Inbox
Summarization**

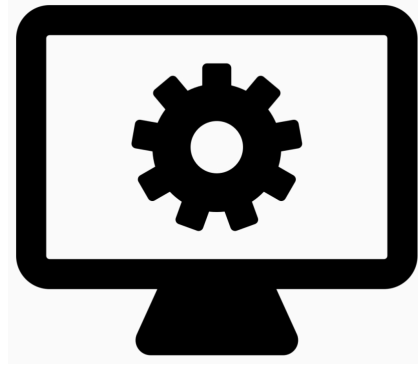


**Legal
Document**



Applications

Software
Documentation



Wikipedia
Articles



Inbox
Summarization

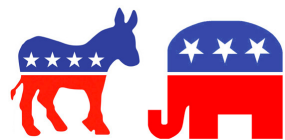


Legal
Document



Overview

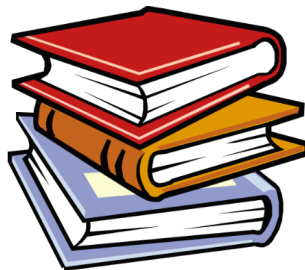
Style



ACL '20
Storytelling '19
ACL '18



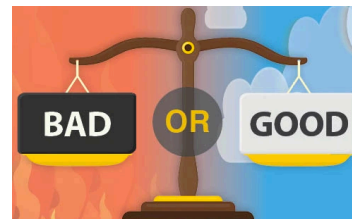
Content



NAACL '19
EMNLP '18



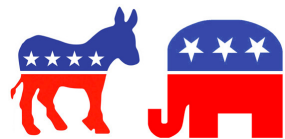
Ethical Considerations



WiNLP '19

Overview

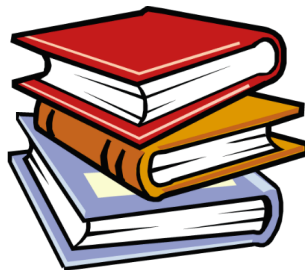
Style



ACL '20
Storytelling '19
ACL '18



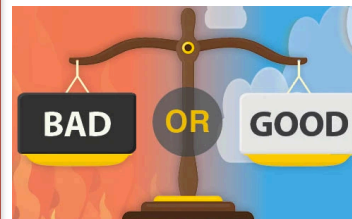
Content



NAACL '19
EMNLP '18



Ethical Considerations



WiNLP '19

Ethical Considerations

- Swear words, obscenity, bias, hate speech
- Broader Impact of controllable text generation
- Social good and bad applications
 - Generate persuasive tweets to spread awareness about climate change
 - Generate persuasive social media content to keep people away from vaccines

Ethical Frameworks

- Understanding ethics and the techniques designed to make systems ethical should be well-grounded in the relevant literature outside AI.

- Generalization principle

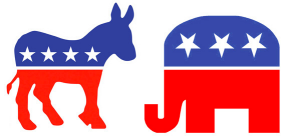
[An ethical decision-maker] must be rational in believing that the reasons for action are consistent with the assumption that everyone with the same reasons will take the same action.

- Utilitarian principle

An action is ethical only if it is not irrational for the agent to believe that no other action results in greater expected utility.

Overview

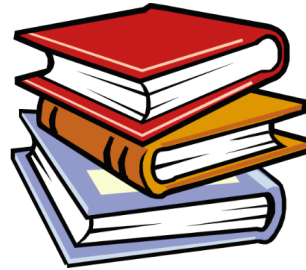
Style



draft ACL '20
Storytelling '19
ACL '18



Content



NAACL '19
EMNLP '18



Ethical Considerations



WiNLP '19



Thank You!

- Alan W Black (co-advisor)
- Ruslan Salakhutdinov (co-advisor)
- Yulia Tsvetkov (committee member)
- Jason Weson (committee member)
- Michel Galley (collaborator)
- Chris Quirk (collaborator)