Controllable Text Generation

Should machines reflect the way humans interact in society?

Shrimai Prabhumoye

Carnegie Mellon University
Language Technologies Institute
We constantly author text!

The key to human-like generation is 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)
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
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.
**Structure**

**The Lead:** The most important information
May include a “hook” (provocative quote or question)

**The Body:** The crucial information
Argument, Controversy, Story, Issue
Evidence, background, details, logic, etc
Quotes, photos, support, dispute, expand

**The Tail:** extra information
Interesting/Related
items like blogs, other editorials
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!
Application

I have a talk to present today!

May the force be with you!

Style
Application
Hey, have you watched Avengers Endgame
Hey, have you watched Avengers Endgame

Yeah, I loved Bradley Cooper in it!
Hey, have you watched Avengers Endgame

Yeah, I loved Bradley Cooper in it!
Hey, have you watched Avengers Endgame?

Yeah, I loved Robert Downey in it!
Application
Application

Hey, have you watched Avengers Endgame
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

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

MT $e \rightarrow f$

encoder  decoder
I thank you, Rep. Visclosky

Architecture

MT e $\rightarrow$ f
encoder decoder

I thank you, Rep. Visclosky

je vous remercie, Rep. Visclosky
I thank you, Rep. Visclosky.

**Architecture**

MT e→f
c
encoder
decoder

je vous remercie, Rep. Visclosky
I thank you, Rep. Visclosky

Architecture
I thank you, Rep. Visclosky

MT $e \rightarrow f$

encoder decoder

je vous remercie, Rep. Visclosky

MT $f \rightarrow e$

coder

$\mathcal{Z}$
I thank you, Rep. Visclosky

Architecture

MT e → f encoder decoder

je vous remercie, Rep. Visclosky

MT f → e encoder

Z
I thank you, Rep. Visclosky

je vous remercie, Rep. Visclosky

Architecture
I thank you, Rep. Visclosky.

MT e → f
encoder  decoder

je vous remercie, Rep. Visclosky

MT f → e
encoder

Architecture

Style 1
decoder

Style 2
decoder
I thank you, Rep. Visclosky

 MT \( e \rightarrow f \)
 encoder decoder

 je vous remercie, Rep. Visclosky

 MT \( f \rightarrow e \)
 encoder

 Style 1
decoder

 I thank you, senator Visclosky

 Style 2
decoder

 I'm praying for you sir.
Train Pipeline
Train Pipeline

\[ \mathcal{Z} \]

- Style 1
  - decoder
  - \( \hat{x}_{\text{style1}} \)

- Style 2
  - decoder
  - \( \hat{x}_{\text{style2}} \)
Train Pipeline

- Style 1
  - decoder
  - $\hat{x}_{style1}$

- Style 2
  - decoder
  - $\hat{x}_{style2}$

- Z
- classifier
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

- Gender
  - Baseline (Shen et al, 2017): 60.4
  - Ours: 57.04
- Political Slant
  - Baseline (Shen et al, 2017): 75.82
  - Ours: 88.01
- Sentiment
  - Baseline (Shen et al, 2017): 80.43
  - Ours: 87.22
Meaning Preservation

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

<table>
<thead>
<tr>
<th>Gender</th>
<th>Political Slant</th>
<th>Sentiment</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Baseline (Shen et al, 2017)</td>
<td>Ours</td>
</tr>
<tr>
<td>0</td>
<td>15.23</td>
<td>43.41</td>
</tr>
<tr>
<td>12.5</td>
<td>14.55</td>
<td>41.36</td>
</tr>
<tr>
<td>25</td>
<td>39.55</td>
<td>45.9</td>
</tr>
<tr>
<td>37.5</td>
<td>35.91</td>
<td>23.18</td>
</tr>
<tr>
<td>50</td>
<td>41.36</td>
<td>40.91</td>
</tr>
</tbody>
</table>

Which transferred sentence is semantically equivalent to the source sentence with an opposite sentiment.
The key to human-like generation is Style, Content and Structure.
Overview

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

- AI assistance deals with *form* (grammar, style, etc.)
- Our goal is to control for *content*
What is our task?
What is our task?
Monkey selfie copyright dispute

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.

Ape-ture priority photographer plays down monkey reports

Chris Cheesman
July 9, 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.

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 animals...
What is our task?

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]
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]
Primary Contribution
Primary Contribution

curated text (context)
Primary Contribution

curated text (context)
Primary Contribution

curated text (context) + document
Primary Contribution

curated text (context) + document =

Primary Contribution

curated text (context) + document = updated text
Primary Contribution

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

Software Documentation

Wikipedia Articles

Inbox Summarization

Legal Document
Data Creation Process
Data Creation Process
Data Creation Process
Data Creation Process

Context

Update
Data Creation Process
Data Creation Process
Data Creation Process

Context

Update
Data Creation Process

Context

Update

HTML News Article
Data Creation Process

Context

Update

HTML News Article

[WikiPedia]

[Common Crawl]
Data Creation Process

Context

Update

HTML News Article

Plain Text of News Article

Common Crawl
Data Creation Process

Total Data Size: 636K

<table>
<thead>
<tr>
<th>News Article</th>
<th>Wikipedia Context</th>
<th>Update</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Models

Generative Models

- Context Agnostic Generative Model (CAG) — Baseline
- Context Informed Generative Model (CIG)
- Context Responsive Generative Model (CRG)

all models have global attention

Extractive Models

- SumBasic
- Context Informed SumBasic
- Oracle

all models are simplistic to infer if context helps in generation
Context Agnostic Model (CAG) - Baseline

News Article

$x_1$ $x_2$ $x_3$

Encoder Vector

$y_1$ $y_2$ $y_3$

Update
Context Informed Model (CIG)

News Article + Wiki Context

Update
Context Receptive Model (CRG)

Encoder Vector

x_1 \rightarrow x_2 \rightarrow \ldots \rightarrow x_n

News Article

Encoder Vector

s_1 \rightarrow s_2 \rightarrow \ldots \rightarrow s_n

Wiki Context

y_1 \rightarrow y_2 \rightarrow y_3

Update

\langle{\text{start}}\rangle

LSTM

LSTM

LSTM

LSTM

LSTM

LSTM

LSTM
## Automated Evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>ROUGE-L</th>
</tr>
</thead>
<tbody>
<tr>
<td>SumBasic</td>
<td>5.6 (5.6-5.7)</td>
</tr>
<tr>
<td>Context Informed SumBasic</td>
<td>7.0 (7.0-7.1)</td>
</tr>
<tr>
<td>Context Agnostic Generative Model</td>
<td>9.1 (9.0-9.2)</td>
</tr>
<tr>
<td>Context Informed Generative Model</td>
<td>16.0 (15.9-16.1)</td>
</tr>
<tr>
<td>Context Receptive Generative Model</td>
<td>14.7 (14.6-14.8)</td>
</tr>
<tr>
<td>Oracle</td>
<td>28.8 (28.7-29.0)</td>
</tr>
</tbody>
</table>

* METEOR and BLEU numbers are consistent with ROUGE-L
Relative Human Evaluation

Which system output is closest in meaning to the reference update?

- CAG
- CIG
- neither

Which system output is more accurate relative to the background information given in the snippet of the article?

- Close to reference
- Coherent to Content

Which system output is more accurate relative to the background information given in the snippet of the article?
Absolute Quality Evaluation

- Grammar: CAG 2.6, CIG 4.3
- Non-redundancy: CAG 1.8, CIG 3.9
- Referential Quality: CAG 2.7, CIG 3.6
- Focus: CAG 2.6, CIG 3.5
- Structure and Coherence: CAG 2.4, CIG 3.2

Scale: (very poor) 1, (very good) 5
The key to human-like generation is Style, Content and Structure
The key to human-like generation is Style, Content and Structure.
Sentence Ordering Task
Ironman builds a time machine to save the world.

Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Thor decapitates Thanos.

Five years later, AntMan escapes from the quantum realm.

Ironman builds a time machine to save the world.
Ironman builds a time machine to save the world.

Ironman steals the Infinity Stones back from Thanos and uses them to disintegrate Thanos and his army, at the cost of his life.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Thor decapitates Thanos.

Five years later, AntMan escapes from the quantum realm.

Hulk travels to New York City in 2012 and convinces the Ancient One to give him the Time Stone.

Ironman builds a time machine to save the world.

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.

Five years later, AntMan escapes from the quantum realm.
Methodology

- Constraint Solving Problem vs Sequence Prediction Task
- For a document with $n$ sentences ($\{s_1 \ldots s_n\}$)
  - $|C| = \binom{n}{2}$ constraints
- Predicted constraints of the form $s_1 < s_2$
- 4 sentences in a document then 6 constraints
  - $\{s_1 < s_2, s_1 < s_3, s_1 < s_4, s_2 < s_3, s_2 < s_4, s_3 < s_4\}$
- Topological sort to find an order given $C$
  - Graph: $s_1 \rightarrow s_2$ if $s_1 < s_2$
Constraint Learning

- **BERT based Representation (B-TSort)**
  - Next Sentence Prediction
  - MLP(BERT($s_1$[SEP]$s_2$))

- **LSTM based Representation (L-TSort)**
  - $h_1 = LSTM(s_1)$; $h_2 = LSTM(s_2)$
  - MLP([h_1; h_2])
Baselines

- **Attention Order Network (AON)**
  - LSTM: sentence representation
  - Transformer: document representation
  - LSTM decoder: generate order

- **BERT Attention Order Network (B-AON)**
  - BERT: sentence representation
Results for NIPS abstracts

<table>
<thead>
<tr>
<th>Metric</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>32.59</td>
<td></td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>61.48</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>76.29</td>
<td>83.45</td>
</tr>
</tbody>
</table>

entire sequence was correctly predicted
Results for NIPS abstracts

- **Perfect Match**
  - AON: 16.25
  - L-TSort: 19.9
  - B-AON: 32.59
  - B-TSort: 43.08

- **Sentence Acc**
  - AON: 50.5
  - L-TSort: 43.08
  - B-AON: 55.23
  - B-TSort: 61.48

- **Kendall Tau**
  - AON: 67
  - L-TSort: 64
  - B-AON: 73
  - B-TSort: 81

- **Rouge-S**
  - AON: 80.97
  - L-TSort: 80.08
  - B-AON: 83.65
  - B-TSort: 87.97

- **LCS**
  - AON: 74.38
  - L-TSort: 71.11
  - B-AON: 76.29
  - B-TSort: 83.45

Absolute position was correctly predicted.
Results for NIPS abstracts

- Perfect Match
- Sentence Acc
- Kendall Tau
- Rouge-S
- LCS

<table>
<thead>
<tr>
<th>Method</th>
<th>Perfect Match</th>
<th>Sentence Acc</th>
<th>Kendall Tau</th>
<th>Rouge-S</th>
<th>LCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>AON</td>
<td>16.25</td>
<td>50.5</td>
<td>55.23</td>
<td>80.97</td>
<td>74.38</td>
</tr>
<tr>
<td>L-TSort</td>
<td>12.19</td>
<td>43.08</td>
<td>61.48</td>
<td>80.08</td>
<td>71.11</td>
</tr>
<tr>
<td>B-AON</td>
<td>32.59</td>
<td>67</td>
<td>64</td>
<td>87.97</td>
<td>71.11</td>
</tr>
<tr>
<td>B-TSort</td>
<td>83.45</td>
<td>76.29</td>
<td>83.65</td>
<td>80.97</td>
<td>76.29</td>
</tr>
</tbody>
</table>

number of inversions to reach correct order
Results for NIPS abstracts

- **AON**
- **L-TSort**
- **B-AON**
- **B-TSort**

Pairs with correct relative order:

<table>
<thead>
<tr>
<th>Measure</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>32.59</td>
<td></td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>55.23</td>
<td>55.23</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>76.29</td>
<td>83.45</td>
</tr>
</tbody>
</table>
Results for NIPS abstracts

- AON
- L-TSort
- B-AON
- B-TSort

- Perfect Match
- Sentence Acc
- Kendall Tau
- Rouge-S
- LCS

- longest common subsequence
### Results for NIPS abstracts

<table>
<thead>
<tr>
<th>Measure</th>
<th>AON</th>
<th>L-TSort</th>
<th>B-AON</th>
<th>B-TSort</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perfect Match</td>
<td>16.25</td>
<td>12.19</td>
<td>32.59</td>
<td>19.9</td>
</tr>
<tr>
<td>Sentence Acc</td>
<td>50.5</td>
<td>43.08</td>
<td>61.48</td>
<td>67</td>
</tr>
<tr>
<td>Kendall Tau</td>
<td>67</td>
<td>64</td>
<td>73</td>
<td>81</td>
</tr>
<tr>
<td>Rouge-S</td>
<td>80.97</td>
<td>80.08</td>
<td>83.65</td>
<td>87.97</td>
</tr>
<tr>
<td>LCS</td>
<td>74.38</td>
<td>71.11</td>
<td>76.29</td>
<td>83.45</td>
</tr>
</tbody>
</table>
Results for NIPS abstracts

- L-TSort performs close to AON
• B-AON performs better than AON
Results for NIPS abstracts

- B-TSort performs the best
Results for NIPS abstracts

- **Perfect Match**
  - AON: 16.25
  - L-TSort: 12.19
  - B-AON: 32.59
  - B-TSort: 32.59

- **Sentence Acc**
  - AON: 50.5
  - L-TSort: 43.08
  - B-AON: 55.23
  - B-TSort: 61.48

- **Kendall Tau**
  - AON: 67
  - L-TSort: 64
  - B-AON: 73
  - B-TSort: 81

- **Rouge-S**
  - AON: 80.97
  - L-TSort: 80.08
  - B-AON: 83.65
  - B-TSort: 87.97

- **LCS**
  - AON: 74.38
  - L-TSort: 71.11
  - B-AON: 76.29
  - B-TSort: 83.45
Results for Human Evaluation

B-TSort vs B-AON
- B-TSort: 41
- No Preference: 28
- B-AON: 31

B-TSort vs Gold
- B-TSort: 26
- No Preference: 20
- Gold: 54

B-AON vs Gold
- B-AON: 22
- No Preference: 24
- Gold: 54
Results

● B-TSort performs best in all metrics for SIND captions, NSF abstract, AAN abstract datasets

● Analysis of displaced sentences
  ● distance between the predicted position and the actual position of a sentence.
  ● B-TSort displaces less sentences.

● Analysis of documents with more than 10 sentences
  ● B-TSort also performs better for longer sentences

● Percentage of mismatch in input and output for AON
  ● Order generation can have a mismatch between the input sentences and the generated orders.
  ● This problem does not arise for our models.
The key to human-like generation is Style, Content and Structure
The key to human-like generation is Style, Content and Structure
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
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.

- Interactional Fairness

- Respect for Autonomy

An agent should not adopt an action plan that the agent is rationally constrained to believe is inconsistent with an ethical action plan of another agent, without informed consent.
The key to human-like generation is Style, Content and Structure.
The key to human-like generation is Style, Content and Structure

**Style**: new model for style transfer and new styles like politeness, political slant, and gender.

**Content**: define new tasks to perform content grounded generation from unstructured data.

**Structure**: new framing of the sentence ordering task and the model is the new state-of-the-art.
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)