



Text-Style Transfer of YouTubers



By: Jonah Bostrom, Jack Bressett, Ben Hayum, & Marwan Lloyd

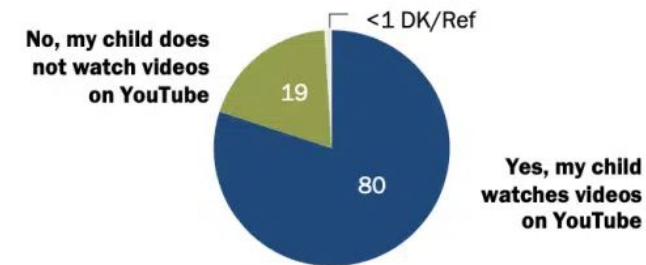
Youtube is the new TV

- YouTube has 2.5 Billion Monthly Users¹
- It is the 2nd most visited website in the world²
- In 2019, 500 hours of footage were uploaded every minute²
- In 2019, 1/3 of American children 8-12 wanted to become Youtubers³

Most parents of young children say their child watches videos on YouTube ...

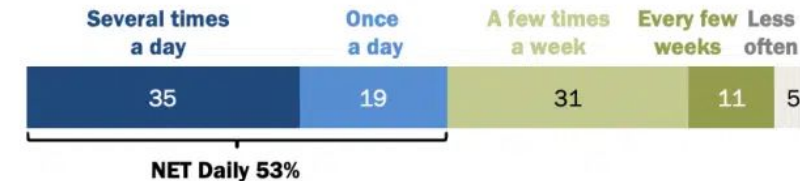
2

% of U.S. parents of a child age 11 or younger who say ...



... and among this group, roughly half say their child does this on a daily basis

Among parents of a child age 11 or younger who say their child watches videos on YouTube, % who say their child does so ...



Note: If parent has multiple children, they were asked to focus on one child when answering this question. Those who did not give an answer are not shown.

Source: Survey of U.S. adults conducted March 2-15, 2020.

"Parenting Children in the Age of Screens"

PEW RESEARCH CENTER



What is Text Style Transfer and why is it useful?

—



Same substance, different presentation

4

Style Attribute	Source Attribute / Sentence	Target Attribute / Sentence
Politeness	Polite: “ Could you please send me the data?”	Impolite: “send me the data!!”
Toxicity	Offensive: “I hope they pay out the ***.”	Non-offensive: “I hope they pay what they deserve. ”
Simplicity	Expert: “Many cause dyspnea , pleuritic chest pain, or both.”	Layman: “The most common symptoms, regardless of the type of fluid in the pleural space or its cause, are shortness of breath and chest pain.”
Biasedness	Biased: “A new downtown is being developed which will bring back... ”	Neutral: “A new downtown is being developed which its promoters hope will bring back...”
Authorship	Shakespearean: “My lord, the queen would speak with you, and presently. ”	Contemporary: “My lord, the queen wants to speak with you right away. ”



Data Overview & Collection





Data Collection & Cleaning

- Used Scrapetube and Youtube_Transcripts API to gather transcripts w/out paying for YouTube API
- Validated manually submitted transcripts via Youtube's official API
- Created CSV's for each Youtuber channel
- Cleaned + Tokenized with NLTK tokenizers
- Created Vocabulary and corpora
- Trained Word2Vec to make relevant Embedding Matrix

Collection Results:

- Significant Data Collected
 - Pro- better, relevant data
 - Con- overwhelming runtime for large YouTube channels

Youtuber	Number of Videos	Total Tokens	Average Tokens Per Video
3Blue1Brown	119	22,190	186.47
Kings and Generals	735	91,230	124.12
ElectroBOOM	198	11,788	59.54
Historia Civilis	83	16,561	199.53
PBS Eons	225	13,649	60.66
Moth Light Media	101	199	1.97
Dr Dray	2,428	676,123	278.47

←
Downsampled
during training



**How can we make a machine that
performs this task?**

—

Model Overview

- During model exploration & lit review we reviewed many projects before finding [Disentangled Representation Learning for Non-Parallel Text Style Transfer](#)⁵

Approach overview used in the paper as well as our project⁵

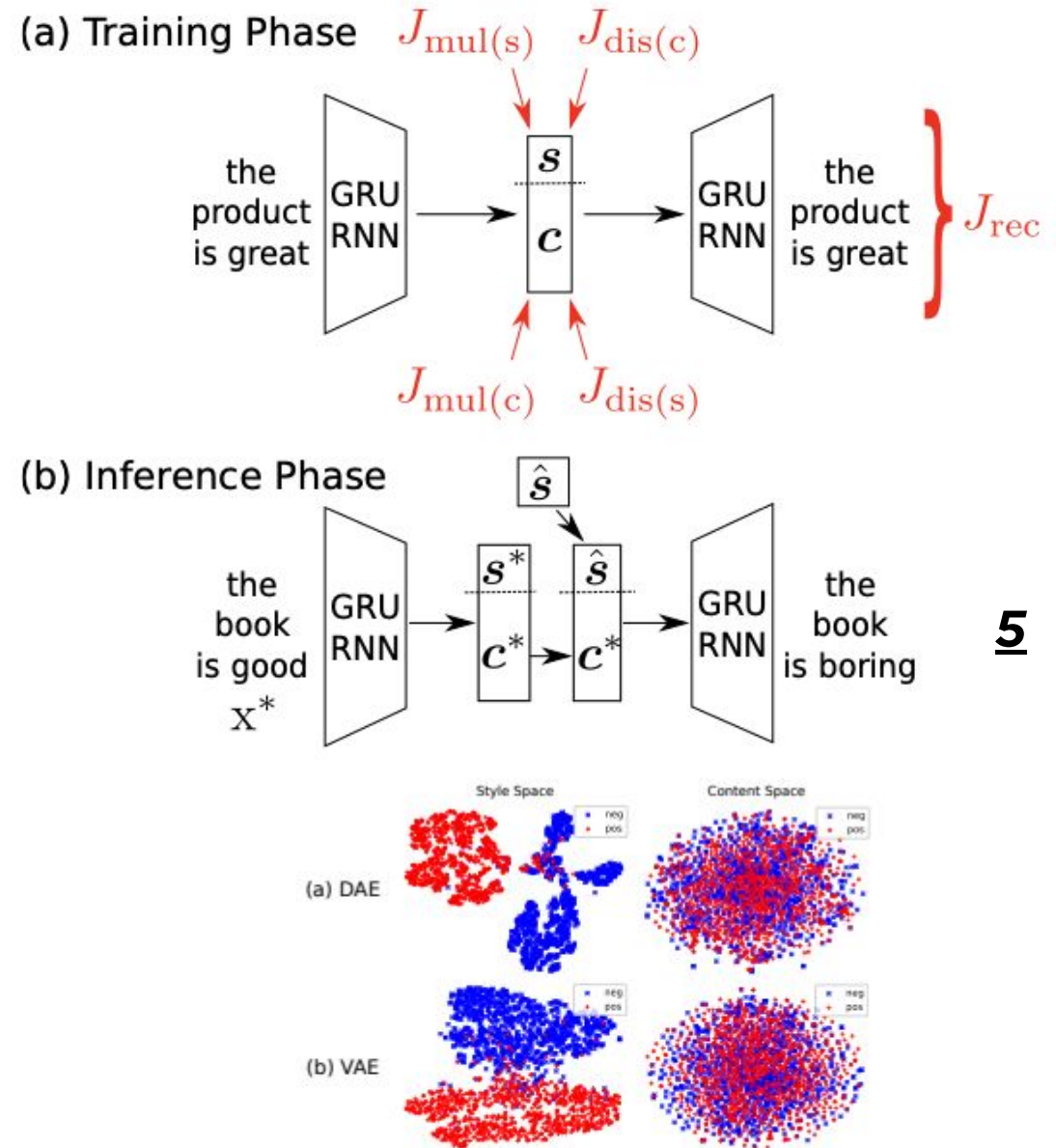


Figure 2: t-SNE plots of the disentangled style and content spaces on Yelp (with all auxiliary losses).



Model Primary Components: Encoder - Adversary - Classifier - Decoder

- Encoder → [Style Embedding, Content Embedding]
 - LOSS: KL Divergence in each latent space → regularize to avoid overfitting and make each latent space represent a standard Gaussian
- Adversary: (Discriminate Style from Content)
 - Style Embedding → Adversary Content
 - Content Embedding → Adversary Style
 - LOSS: Get style information out of content latent space & vice versa
- Classifier: (Reward/Punish Prediction of Author)
 - Style Embedding → Label → Reward/Punishment
 - Content Embedding → Label → Reward/Punishment
 - LOSS: Correctly identify style and content from latent space
- Decoder → Reconstructed Sentence Input
 - LOSS: Reconstruction Loss → Build back original sentence

Middle of Model



Modelling Issues & Results





Major Problems Impeded Final Results

- Size:
 - Vocabulary Size – Limit 10,000 of 45,000 total vocab size
 - Sentence Size – Limit 20 when mean is 50 word tokens
 - Word Embedding Size – Limit 64 when paper is 100-300 dimensions
 - Unique Model Architecture – Forced to store large intermediary tensors (the paper had donated powerful GPUs)
 - Best model train session 6 epochs until 35 GB memory limit hit
- Data:
 - Numerous decisions when cleaning
 - Over-inflation of padding created model overfitting
 - Data collection and tokenization process not fully reliable (quality + quantity)



Sentence Generation Progress

Iteration	Original sentence	Style Transferred Sentence: K&G (History)	Style Transferred Sentence: 3B1B (Mathematics)
1	"the kings and generals channel makes videos about this topic"	<pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>	<pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad> <pad>
2	"the kings and generals channel makes videos about this topic"	frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline frontline	supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting supporting
3	"the kings and generals channel makes videos about this topic"	<pad> <sos> reducing acts , 25 saints unchallenged velocity brains resentment 2/16th continuously creature 4,000 kütahı kütahı etc sick independently	<pad> <sos> reducing acts 25 brains , saints saints velocity velocity 2/16th sick kütahı continuously resentment helped rapid harsh etc



Wider Implications





Legal and Ethical Consequences

- Intellectual Property and Authorship
 - If a youtuber spends hours or days making a youtube video that another can easily style transfer with the click of a button, who is the true owner of the video? The youtuber, the other person, or the model?
 - Who gets paid the proceeds? What will the copyright laws look like?
- Misinformation and Weakening of Democracy's Epistemics
 - Falsified videos can also be used to generate inflammatory content intended to radicalize populations, slander public figures, or incite political instability.
 - May also lead to the "Liar's Divided" where guilty individuals may be able to successfully deny the authenticity of genuine content.
 - Elections may be swayed. Whoever has better Generative AI and ability to A/B test messages will be the winners of elections from 2028 onwards.

Regulations

External:

- European Union AI Act, 2021 ⁶
 - Legislation that strict rules for AI systems that are considered high-risk, such as those used in critical infrastructure or that can affect individuals' fundamental rights
- United States, 2019 ⁷
 - Representative Yvette Clarke introduced a bill to mitigate risks from deepfakes that required "mandatory labeling, watermarking, or audio disclosures for all 'advanced technological false personation records'"
 - Insufficient support for 1st amendment reasons

Internal:

- No Model open sourcing
- Youtuber vetting process for who can be used in the model
- Give youtubers the option to opt in to being used by these models or not
- Watermark content with easily identifiable marks and not allow the option for these to be removed



Conclusion

- Our working results demonstrate the learning task is possible
- Better cleaning will be necessary for larger scale
- Better training resources required (Colab Pro not sufficient)
- Equally important, many legal & ethical areas will have to be addressed as the technology continues to develop



Works Cited

1. [23 YouTube Stats That Matter to Marketers in 2023](#)
2. [Bloomberg.com - Eight Numbers That Show How Big YouTube Has Become](#)
3. [CNBC.com - Forget law school, these kids want to be a YouTube star](#)
4. [An Introduction to Text Style Transfer](#)
5. [Disentangled Representation Learning for Non-Parallel Text Style Transfer](#)
6. [Axios.com - Scoop: Schumer lays groundwork for Congress to regulate AI](#)
7. [EFF.org - Congress Should Not Rush to Regulate Deepfakes](#)