

Topic-aware Contextualized Transformers: Appendix

A OVERALL ALGORITHM AND OPTIMIZATION

To infer topic-aware contextualized Transformers, we consider a hybrid of stochastic gradient MCMC, used for the PGBN topics ϕ_m^t , and auto-encoding variational inference, used for the parameters of both the inference network (encoder) and Transformers. Here, we provide the overall algorithm of our contextualized Transformers in Algorithm 1, where we take the GPT-2 as an example.

Algorithm 1 Stochastic-gradient MCMC and autoencoding variational inference for contextualized Transformers

Set mini-batch size as N , the number of topic layers T , the topic numbers M_1, \dots, M_T , and the segment length I ;
Initialize encoder parameters \mathbf{W}_I ; Initialize topic model parameters $\{\Phi^t\}_{t=1}^T$; Initialize language model parameters Ω from pre-trained GPT-2.
for $iter = 1, 2, \dots$ **do**
 Randomly select a mini-batch of N segments consisting of I tokens to form a subset $\{\mathbf{u}_1, \dots, \mathbf{u}_i, \dots, \mathbf{u}_I\}_{n=1}^N$;
 Add the token embedding to each token $\left\{ \{\mathbf{E}_{u_i} = W\mathbf{E}_{u_i} + P\mathbf{E}_{u_i} + T\mathbf{E}_{u_i} + E_B\}_{i=1}^I \right\}_{n=1}^N$;
 Concatenate the segment embedding to each segment $\{[SE^1 + E_A, \dots, SE^T + E_A, \mathbf{E}_{u_1}, \dots, \mathbf{E}_{u_I}]\}_{n=1}^N$;
 Input contextualized embeddings of N segments into transformers;
 for $layer = 1, 2, \dots$ **do**
 Transform through residual-connected masked multi-head attention and layer normalization;
 if integrate topic attention with M_t topics at layer t **then**
 Calculate $\tilde{\Phi}^t$ by topic self-attention;
 Transform through residual-connected topic select-attention and layer normalization;
 end if
 Transform through residual-connected feed forward network and layer normalization;
 end for
 Calculate $\nabla_{\Omega} (L_{LM})$ according to (1) and update Ω ; Calculate $\nabla_{\mathbf{W}_I} (L_{LM} + L_{TM})$ according to (1) and (2) and update \mathbf{W}_I ; Sample θ^t via \mathbf{W}_I , described in (3), with which to update $\{\Phi^t\}_{t=1}^T$ according to (6);
end for

To update the global parameters $\{\Phi^t\}_{t=1}^T$ of PGBN an local representations of segment or token $\theta_{1:N}^{1:T}$ in detail and give a complete one in Algorithm in 1.

$$L_{LM} = - \sum_i [\log P(u_i | u_{<i}; \Omega, \{\Phi^t, \theta^t, \tilde{\theta}^t\}_{t=1}^T) - L_{i,sparse} - L_{i,predict}], \quad (1)$$

$$L_{TM} = - \sum_{n=1}^N \mathbb{E}_Q [\ln P(d_n | \Phi^1 \theta_n^1)] + \sum_{n=1}^N \sum_{t=1}^T \mathbb{E}_Q \left[\ln \frac{q(\theta_n^t | d_n)}{P(\theta_n^t | \Phi^{t+1} \theta_n^{t+1}, \tau_n^{t+1})} \right], \quad (2)$$

Sample the multi-layer representations $\theta_{1:N}^{1:T}$ for all segments: All segments of the target corpus are treated as BoW vectors (d_1, \dots, d_N) , ignoring word order. We introduce a Weibull hybrid autoencoding inference (WHAI) network (encoder) Zhang et al. (2018) for PGBN (decoder). Denoting $Q = \prod_{t=1}^T \prod_{n=1}^N q(\theta_n^t | d_n)$, the negative ELBO of PGBN be expressed as (2). Similar to Zhang et al. (2018), we define $q(\theta_n^t | d_n) = \text{Weibull}(\kappa_n^t, \lambda_n^t)$, a random sample from which can be obtained by transforming standard uniform noises ϵ_n^t as

$$\theta_n^t = \lambda_n^t (-\ln(1 - \epsilon_n^t))^{1/\kappa_n^t}. \quad (3)$$

where \mathbf{k}_n^t and $\lambda_n^{(t)}$ are the parameters of θ_n^t that are nonlinearly transformed from the hidden units \mathbf{h}_n^t as

$$\mathbf{k}_n^{(t)} = \ln[1 + \exp(\mathbf{W}_{hk}^{(t)} \mathbf{h}_n^{(t)} + \mathbf{b}_1^{(t)})], \quad (4)$$

$$\lambda_n^{(t)} = \ln[1 + \exp(\mathbf{W}_{h\lambda}^{(t)} \mathbf{h}_n^{(t)} + \mathbf{b}_2^{(t)})], \quad (5)$$

where $\mathbf{h}_n^{(t)}$ are deterministically nonlinearly transformed from \mathbf{d}_n . All parameters in encoder network be denoted as \mathbf{W}_I , which can be learned via SGD with negative ELBO expressed in (2).

Sample the hierarchical connection weight matrices $\{\Phi^t\}_{t=1}^T$: For ϕ_m^t , the m th column of the loading matrix Φ^t of layer t , its sampling can be efficiently realized as

$$(\phi_m^t)_{i+1} = \left[(\phi_m^t)_i + \frac{\varepsilon_i}{P_m^t} \left[(\rho \tilde{\mathbf{A}}_{:,m}^t + \eta_0^t) - (\rho \tilde{A}_{:,m}^t + M_{t-1} \eta_0^t) (\phi_m^t)_i \right] + \mathcal{N} \left(0, \frac{2\varepsilon_n}{P_m^t} [\text{diag}(\phi_m^t)_i - (\phi_m^t)_i (\phi_m^t)_i^T] \right) \right]_{\angle}, \quad (6)$$

where $[\cdot]_{\angle}$ denotes the simplex constraint that $\phi_{m,k}^{(l)} \geq 0$ and $\sum_{m=1}^{M_t} \phi_{m,k}^{(t)} = 1$, P_m^t is calculated using the estimated FIM, ε_i denotes the learning rate at the i th iteration, both $\tilde{\mathbf{A}}_{:,m}^t$ and $\tilde{A}_{:,m}^t$ come from the augmented latent counts A^t , and η_0^t denote the prior of ϕ_m^t . More details of TLASGR-MCMC can be found in Cong et al. (2017).

B VARIETIES OF TOPIC-AWARE CONTEXTUALIZED TRANSFORMERS

To further evaluate the efficiency of varieties of our proposed model, we change the layer number of topic model and the intervention types from topic model to GPT-2 on PTB dataset. The results are listed in Table 1, where TM l denotes the l -th layer of PGBN, LM l denotes the l -th layer of GPT-2. "TM1 with LM1-4" represents that we integrate the topic embeddings of the 1-th topic layer into the 1-th, 2-th, 3-th, and 4-th layer of GPT-2. As we can see, multi-layer topics provide more performance gains than shallow layer topic models. Besides, intervention with equal space from topic models to transformer performs best results in both shallow and deep models, which is accordance with the conclusion in Rae and Razavi (2020).

Table 1: Perplexity of varieties of topic-aware contextualized GPT-2 on PTB dataset.

Model	intervention	# Param	PTB
baseline + fine-tune	None	117M	15.22
+ 1-layer-Topic model (100)	TM1 with LM1-12	117+5.22M	15.15
+ 1-layer-Topic model (100)	TM1 with LM1; TM1 with LM4; TM1 with LM8	117+5.12M	15.12
+ 1-layer-Topic model (100)	TM1 with LM4; TM1 with LM8; TM1 with LM12	117+5.12M	15.10
+ 2-layer-Topic model (100,80)	TM1 with LM1; TM1 with LM4; TM2 with LM8	117+5.19M	15.10
+ 2-layer-Topic model (100,80)	TM1 with LM4; TM1 with LM8; TM2 with LM12	117+5.19M	15.07
+ 3-layer-Topic model (100,80,50)	TM1 with LM1-4; TM2 with LM5-8; TM3 with LM9-12	117+5.34M	15.00
+ 3-layer-Topic model (100,80,50)	TM1 with LM1; TM2 with LM4; TM3 with LM8	117+5.24M	15.00
+ 3-layer-Topic model (100,80,50)	TM1 with LM4; TM2 with LM8; TM3 with LM12	117+5.24M	14.92

C TOKEN EMBEDDING WITHIN DIFFERENT CONTEXTS

To verify the efficiency of contextual token embedding, we visualize the embedding learned with our method, where we take the token "right" in two different sentences from the WT2 validation set as example. In Fig.1 (a)-(b), we show the nearest neighbors of "right" in two sentences based on cosine similarity respectively, by mapping their embedding to the three-dimensional subspace with PCA. As shown in Fig.1 (a), the nearest neighbors of "right" are usually semantically related with the words including "Legal, Liberties, moral, own" and others, which is reasonable since its corresponding sentence is about freedom. We also visualize the embedding and neighbors of "right"

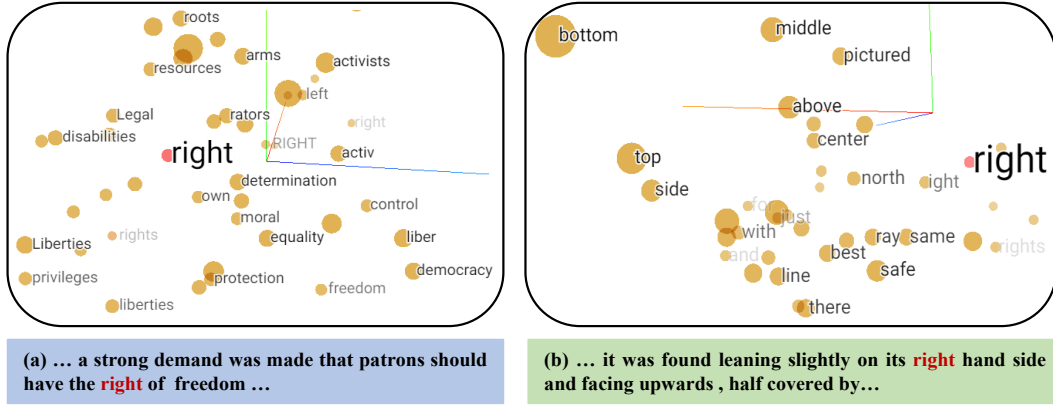


Figure 1: Illustration of the neighboring tokens of "right" in different contexts at the embedding space

given another context in Fig.1 (b), whose neighbors are mostly about “position”. This suggests that the our proposed token embedding can capture the contextual information of a token directly in input space, which can be attributed to the global semantics from the target corpus and localized representation of a token given its context. This is different from standard Transformer taking the static word embedding as input, which allows different internal representations of a token depending on its context using the attention mechanism, learning only a segment-level contextualization.

D ATTENTION VISUALIZATION

We provide visualization of the attention between virtual words and real words learned by the GPT-2 model on the PTB validation set. As the topic number of three-layer PGBN is set to [100, 80, 50], each segment shares 230 virtual words as contextualized embedding vector placed before the word sequence of the current segment. As shown in Fig. 2, we visualize the attention between 230 virtual words and the first 230 real words in the current segment from the bottom layer to the top layer.

As we can see, the attentions between real tokens and virtual tokens are dense at the bottom layers, which have a wide attention distribution over the entire topics. When moving upward, some tokens tend to focus on fewer topics and become more sparse, which may contribute to more contextual information of higher layers. This is intuitive that representations at lower-level layers need to screen the entire virtual topics to decide where to focus for higher-level layers. We see an almost uniform attention over the entire virtual topics of some tokens, which is quite intuitive, as some tokens have fuzzy topic information and couldn’t attend to specific topics.

As for the attention distribution within real words, the overall trend is to focus more on the nearby preceding tokens than the far away ones. However, it is also very clear that the wider attention distributions over entire preceding context are shown in higher layers, which well fits our intuition that the contextual information accumulates as the layers add up.

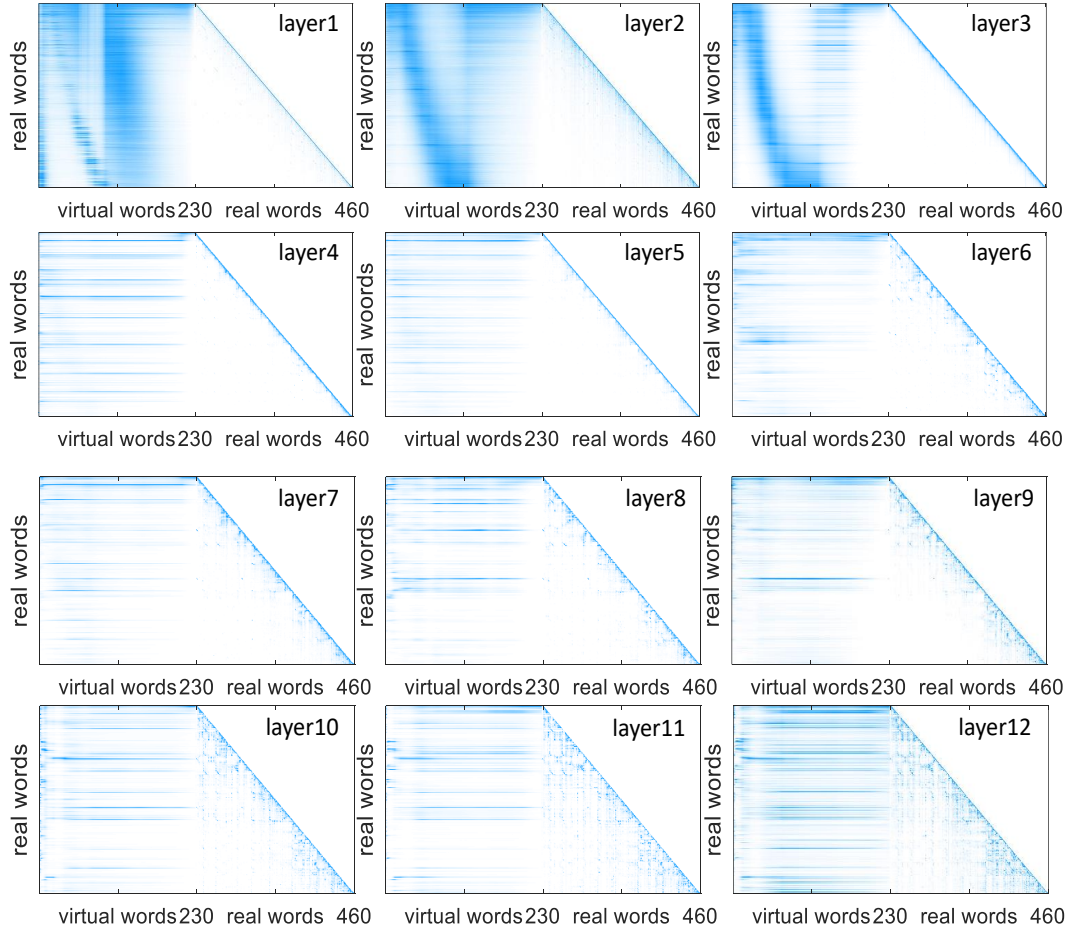


Figure 2: Visualization of the attention between virtual words and real words at layers 1-12 of GPT-2. In each subplot, each row corresponds to a context (real word) location and each column corresponds to a target (virtual/real word) location.

E TEXT GENERATION GUIDED BY TOPICS

Given the learned contextualized Transformers, we can sample the sentences guided by a combination of the topics from different layers. Shown in Fig. 3, we give three semantically related topics obtained with $\{\Phi^t\}_{t=1}^T$. Note that we do not perform any cherry picking and present the first three examples we generate in the paper. As we can see, the generated sentences conditioned on the combination of topics are highly related to the given topics in terms of their semantical meanings but not necessarily in key words, indicating the transformers are successfully guided by contextualized topic information. These observations suggest that contextualized Transformers have successfully captured syntax and global semantics simultaneously for natural language generation.

Topics: <ul style="list-style-type: none">● layer 1: murder assassination bury evil hell certified● layer 2: attack forces gun battle fighting killed wounding● layer 3: death died unknow hell wrong character
➤ In January 2015, Omar Mateen allegedly stabbed 14 women in the back as they were shopping in a New York neighborhood and left them to fend for themselves. He was arrested on a minor misdemeanor charge of murder after his killing spree.
➤ By 9 p.m. Sunday morning, the officers, who were responding to a call about an armed man, had found the SUV. But after their encounter, Ms. Pichon was handcuffed to a police vehicle and placed in a holding cell, where he was given medical attention. On Monday, Mr. Pritt said a lawyer representing his client did not respond to requests to speak to the media.
➤ On the case of Michael Darnell, 35, who was shot Wednesday at the Crossroads Grill, he said it would be a "game on." He said the officer said he could find no evidence of marijuana around the block.

Topics: <ul style="list-style-type: none">● layer 1: play manager joined finished signed hard coast started● layer 2: university game career member manager joined finished● layer 3: league career member coast manager signed started
➤ Thanks to you, we now have over 100k people playing on our mobile game, and our first round of digital subscriptions means we are ready to ship this Kickstarter! After having already done a few updates with our new online backers who got into the game on our Kickstarter (some who still get an early access to the games), I am truly, truly excited to get back on the road with our new people.
➤ It takes a special kind of person to play the game! Every character is unique, every weapon has incredible potential, a small community of fans are ready to help us out, and every character has its own strengths and weaknesses. I am going to be talking about all of that in more detail with you soon when more features are released.
➤ Now that we have many more exciting game features, and even more game features to share with the community, it is time to start trying to sell on a few of them, too!

Figure 3: Generated sequences guided by a combination of topics.

F TEXT GENERATION GIVEN PRECEDING CONTEXT

We present some generated text based on the fine-tuned GPT-2 model. The preceding context mainly focus on the hot news reports, and three generated examples given preceding context are shown in Fig. 4. Note that we do not perform any cherry picking and present the first three examples we generate in the paper. We also show the ability of our model to generate long paragraphs from preceding context, as shown in Fig. 5 and 6. It is clear, our fine-tuned model based on GPT-2 is powerful for generating long text paragraphs, whose generated text are highly semantic-related to the preceding context. Note that with the topic information extracted from preceding context, our proposed model can generate both the semantically-meaningful words stay on the same topic, and the novel content which may not exist in the original corpus. As shown in Fig. 7 and Fig. 8, we input the same preceding context and compare the generated context of our proposed topic-aware contextualized GPT-2 and the original pre-trained GPT-2. We find that our model is able to remember the semantic information with longer dependencies beyond current segment, and generate semantically consistent paragraphs.

<p>● Preceding context : black lives matters</p>
<p>➤ . If we are to move beyond the concept of hate towards racism and the idea of free speech, then it must be clear that every individual – and even those of us who hate, but don't, can hate and that all forms of hate are destructive (though we can go further with others we disagree with) we must stop them from happening – and we must demand that those who we hate – do something to stop them.</p>
<p>➤ and in particular, lives of children who have endured brutal attacks. But the question of when this kind of violence is necessary to justify a person's violence still has the effect of making people feel powerless or the police complicit.</p>
<p>➤ and that's what the United States is for, we support the people who are fighting for them – the people who are in the fight for the future of the United States," he said at the event. "We're not in any war in this country. Because there is no war in this country to destroy the United States, who is in this fight are people who are defending it, are defending it – all of these people. And that has absolutely nothing to do with what we stand for."</p>
<p>● Preceding context : The number of deaths from the coronavirus pandemic exceeds 100,000 in the United States, according to an analysis of data from the Centers for Disease Control and Prevention.</p>
<p>➤ More than 8,000 of his colleagues and family members have been infected with a virus that was recently reported to have died in the area. The latest CDC data is the most likely to point to the pandemic. It is especially devastating for the most vulnerable people who need health care and that are at risk for transmission, including those in a vulnerable group in the United States. But the virus was also transmitted to infected people in the first place and to those still affected in an outbreak that has killed as many as 20,000. As infected people who survive the outbreak will risk the life of thousands, the results might prove devastating to those affected.</p>
<p>➤ "This outbreak is, at a time of high epidemic demand, one of the most intense and most contagious outbreaks for public health and health safety worldwide. The public outbreak is highly contagious, not only in the USA but globally," CDC. In the United States, the disease, spreading with an increasing intensity nationwide, has infected nearly 20 countries.</p>
<p>➤ "The number of deaths from the coronavirus outbreak can 't tell who might have been affected," said Dr Paul J. Kennedy, an epidemiologist from Washington University at St. Louis University and the National Institute for Healthcare Epidemiology, in the statement. "The results of the CDC study are a new low that is too many to contain." Katherine M. Gartan, the vice president for immunization at the CDC, said that CDC reports on this illness are in poor health. "This is a big concern," she said, "and we don't know where the epidemic is. To do it right we have to act faster, the CDC report tells us, as we should. The more they report, the more likely they go out to work on it."</p>

Figure 4: Generated sequences given preceding context.

● **Preceding context :** The World Reopens, Despite Skyrocketing Coronavirus Cases. The number of infections is rising faster than ever, but many countries have decided that this is the moment to ease lockdown restrictions. Two months ago, when there were roughly one million confirmed coronavirus cases and the primal politics of survival was sweeping the world, shutting down was the order of the day. This week, the number of cases soared past seven million, with 136,000 new infections detected on Sunday alone, the highest single-day total since the pandemic began. The order of the day? Reopening. Terrified after watching economies built over the course of decades hollow out in a matter of weeks, countries seem to be saying, in effect: Enough. For health officials who have been watching the virus with alarm as it began claiming a foothold in continent after continent, it is a dizzying moment.

● **Generated paragraph:** The World Reopens, Despite Skyrocketing Coronavirus Cases. The number of infections is rising faster than ever, but many countries have decided that this is the moment to ease lockdown restrictions. Two months ago, when there were roughly one million confirmed coronavirus cases and the primal politics of survival was sweeping the world, shutting down was the order of the day. This week, the number of cases soared past seven million, with 136,000 new infections detected on Sunday alone, the highest single-day total since the pandemic began. The order of the day? Reopening. Terrified after watching economies built over the course of decades hollow out in a matter of weeks, countries seem to be saying, in effect: Enough. For health officials who have been watching the virus with alarm as it began claiming a foothold in continent after continent, it is a dizzying moment. But that's what has caused the first wave of outbreaks - on the way to being seen as a "one in 10" epidemic, says one of the most experienced infectious-reversal researchers on the globe. In the week since the deadly virus was discovered, the virus has already found and spread, and in the months since the latest, there have been over 14,000 infections. There may go the number, however, as many are spreading on the ground. It's likely that the numbers have been at least one-quarter of the number expected to see one out of thousands, to say nothing of a second, which is a number that has been on the rise since. It's going to take an emergency, this time around, to quell the epidemic on our home planet. There are now over half a million people living under the threat of a "one in 10" epidemic, which is the third highest, following the latest and most recent pandemic. The situation is "worrisome," said a recent research by leading scientists, the international scientific association for the health of the Earth. The global virus, which became highly contagious about five weeks ago, has made a new all the more acute. It's a fact, said Dr. David Crenshaw, of the Department of Emergency Management in the US, a "huge concern" - and an all-time low, but he says, like Ebola.

Figure 5: Generated paragraph given preceding context.

● **Preceding context :** Public Opinion Rarely Moves Fast, but It Has on Black Lives Matter. In the last two weeks, American voters' support for the Black Lives Matter movement increased almost as much as it had in the preceding two years. American public opinion can sometimes seem stubborn. Voters haven't really changed their views on abortion in 50 years. Donald J. Trump's approval rating among registered voters has fallen within a five-point range for just about every day of his presidency. But the Black lives Matter movement has been an exception from the start. Public opinion on race and criminal justice issues has been steadily moving left since the first protests ignited over the fatal shootings of Trayvon Martin and Michael Brown. And since the death of George Floyd in police custody on May 25, public opinion on race, criminal justice and the Black Lives Matter movement has leaped leftward. Over the last two weeks, support for Black Lives Matter increased by nearly as much as it had over the previous two years, according to data from Civiqs, an online survey research firm. By a 28-point margin, Civiqs finds that a majority of Americans support the movement, up from a 17-point margin before the most recent wave of protests began.

● **Generated paragraph:** More than six in 10 white voters who are white have either a strongly held belief that violence should be a top cause for concern, or view it as a low-hanging fruit in American society. A majority of African-American voters have at least a higher willingness to go to the political polls. But American public opinion is more mixed about social policy and issues in recent weeks. The latest recent protests in Baltimore, as the protests continued, have put many Americans at odds with their political, civic-diversity, and racial background. There has been little change in public opinion since the last protest, as the American public remains sharply divided on most issues, and for the first time in almost two decades of race and the political-politics level of the U.S.S. government. The public attitude has not changed substantially since the recent protests began. At this year's Republican-led Washington protests, about half of the protesters were white. As a result, at least half of the more than 5,000 people who went to the National Mall in protest at the police department shooting of black teenagers and for a series of fatal police shootings have also been black. The protests have a lot of new issues for Americans to confront. They face challenges and in order to help deal with them, we have to put their interests ahead of the interests of the American people. The issues of police-community relations, civil rights, and police-community relations as well as other issues of civic-justice and social issues have created a new divide in American life, and more Americans are getting more, and not less, satisfied with its progress. Black and youth need the same answers about how to respond to racial and, ultimately, criminal justice issues. In the United States as a whole, American political leaders and voters tend to be working-class or, in the extreme case, white. The problems of race and the black social and criminal justice issues are on the public face. In fact, in the United States as a whole—and of more diverse country at and right over the last 60 years after the Second African-American Revolution—the public is not in that class. In the United States as a whole, most voters have a strong interest in black issues, as do many American political leaders, as they have a more liberal American population. But Black voters are also growing more politically and social as well as political political. The national black American population has continued to rise to about 18 percent of US white population, and by this same measure, more a majority of the black American community wants to make the most of the federal government. These new white share of the American political and civil-rights landscape has been a very high rate. When the national black American population has risen to 6-5, black political-rights leaders say at least as many have voted in this presidential election for, or had a voice in black public-relations issues as they say now of economic affairs. This, while not always a sign that American politics has improved as much as in America. The American public opinion in this field has been divided more on the one hand in the past two years. This is partly because Americans are less than their share of the White and Black share of public opinion has changed. The national Pew poll found that in an age-to-population relationship, a majority of Americans has a less clear (though still overwhelming) belief in the political direction of the U.S. Government. The change in American public opinion of this national race—in the last six or seven years, over 80 percent—was by a long way a result of the American public's understanding of the political power of blacks, a demographic that still can and does have a share of the racial class in the country.

Figure 6: Generated paragraph given preceding context.

<p>● Preceding context :</p> <p>This pairing was the first outfit I thought of when I bought the shoes. It's like a summer version of this Jake Grantham outfit; in fact, my shoes are close to the colors of his Nike Racers! Instead of a heavy Harris Tweed jacket and denim shirt, I'm wearing a cotton DB jacket and a linen shirt. Both fabrics (in these colors) are an absolute must for summer, as they go with both dark and light pants! As you can see, they pair wonderfully with the dark jeans and shoes. It's a pseudo menswear/prep outfit. Overall, this is a very casual outfit which is why I paired my sneakers with it. I'm not about wearing a full wool suit with sneakers (as GQ shows a lot) but I'm definitely open to keeping things casual, like this cotton DB. Casual fabrics are key to pulling off your sneakers in a dressed down menswear outfit. I'd even suggest to wear these sneakers with a khaki chino suit or a white linen suit. Just be sure to ditch the tie or wear a tee or polo; wearing a tie with sneakers is a bit too much</p>
<p>● GPT-2:</p> <p>for me!</p> <p>I know that some people might be opposed to wearing sneakers in a turtleneck, but I wanted to be true to myself, so I went with a slim fitting turtleneck in a color more similar to my favorite color of the day. Overall, it's a classic turtleneck, with nothing too flashy. I did swap my shoes for these ASICS Tiger Killshots and I wish I hadn't did that. The shoe itself is very comfortable, however, I found that wearing them in these sneakers made my feet look a bit larger. I'd wear them with denim, chinos, whatever, but would never recommend wearing them alone. There are just too many variables involved.</p>
<p>● Topic-aware contextualized GPT-2:</p> <p>if you want to wear your clothes casual.</p> <p>I look pretty casual now, but the other jeans just don't always match. Here at first, I was pretty picky with my pants, but I'm now really in the top gear. This outfit is so flattering and so sexy for jeans and shorts, when you feel comfortable wearing. This summer outfit is also a good time to wear a skirt. It's actually a more formal shirt if only because it doesn't look like a top and this dress is a good for this kind of thing. And as is your choice. I like jeans and I'm wearing my dress, but the jeans just do not do it justice. One thing I wear now that would not be my current wear is a high street blouse. This outfit would make a great summer dress to outfit my jeans to give me casual outfits. I' t the only real choice when your dress is very dress shirt. My blouse is a great way to get a look I haven't worn for long, when I look over. You wouldn't really want this pants. There is also a high jumper, but the outfit I really like is that with the hat and gloves I have been wearing the entire time. This jacket has the usual button on back at its on the right side of the front for my suit. Now it is a good time to wear the shirt, if you want the look like that of a top. I always wear a paj. So to give my denim outfit another go, here comes all my best of a denim outfit, and I just don't can you get into it. Here you should start wearing this hat so it's that it, with that tie I have a hat as well to give it one good go. For the trousers I had from my shirt, this shirt is my best summer clothing for my body.</p>

Figure 7: Generated paragraphs given preceding context compared to GPT-2 (cite from Radford et al. (2019)).

<p>● Preceding context :</p> <p>Corporal Michael P. Goeldin was an unskilled laborer from Ireland when he enlisted in Company A in November 1860. Goldein survived the war. On 10 June, Artificer William Collins was promoted to corporal, probably to fill a combat leadership void for the crossing of the James River. Collins's service record does not reflect the qualities he demonstrated to earn this promotion, but he had obviously overcome some serious problems. Born in Sacketts Harbor, New York, Collins enlisted in the company in December 1853 at the age of twenty-two, and reenlisted in December 1858. Just a month before the war began in April 1861, Collins went "over the hill" and was not caught until three years later. Returned to the company on 22 March 1864, he was tried</p>
<p>● GPT-2:</p> <p>in June of that year. Sentenced to be shot for desertion, he remained on active duty. Captain James J. Dall was born in Virginia in 1829 during the Revolution and joined his father in the Army in 1836. He served with the 2d Virginia Infantry until 1845, when he went to the Western Reserve Regiment; and, on his discharge, he reenlisted with Company I, 25th New York Volunteers (which in 1845 numbered about 4,500 men). One of his brothers, James, served with his company in the same capacity in the same brigade. While the regiment remained in Connecticut through 1846, Captain Dall served with Company I at various times before his discharge. A veteran of the Mexican War, he had served two years on the staff of Brigadier General John J. Sullivan at San Antonio. During the Peninsula Campaign of 1863, he commanded a company of twenty-three and was in charge of the battery of fifty-four heavy mortars and the gunboat Pinta of the Battery, a gunboat and several other boats. Captain Dall was on active duty with Company I in the fall of 1865. Two years later, he was promoted to second lieutenant.</p>
<p>● Topic-aware contextualized GPT-2:</p> <p>he was tried for treason and later on in the court of the D.J. He was sentenced on 17 April, and was executed on 9 May. He was also sentenced to life imprisonment in the camp at Brandy with an appeal on 10 June, with the company receiving the order to withdraw, on a condition that he would not return to the American frontier. He returned to the company and had five years of his service, and the regiment received one year of enlistment. After retiring to Brandy, he worked as a clerk for his uncle at Wilkes for six years. He was discharged in April of 1864 when he was arrested under the pretence of being a lieutenant of the American Legion but did not serve. He returned to the company, and was also assigned to the Company S, at Ponder Point, Ohio. The regiment received five years of enlistment, served in service the next year that he enlisted, and was returned to the New York Corps in 1884. He was sent to the West Company under the command of Company C in October 1875. As the new Confederate Army took all of its territory with the same General, the men were not able to fully complete his journey until it had already moved down the Mississippi, when he was received on command by General H. W. Davis. With the regiments was enlisted, a number of these on all four regiments in March. During the first day of training there that they were selected to serve on the North, under General H. E. Davis. During that week the American corps, and the Army of Northern Virginia (North Carolina, Army of the United States Confederate Veterans) was divided. By 1877, an active-ment and full-service corps were now the Army of Fort Jackson. In January the army was transferred to the East Company by a special command and the Army of the Northern Virginia (E.W. Hatton. New. Col, West Company. New. Col, and Fort. Jackson. All the other regiments had the same division, which at the end of this year was called the division of the United State Army. In April 1784 the troops was to be sent in to assist on the new battle. Captain C. William H. Brown. First. S. A.</p>

Figure 8: Generated paragraphs given preceding context compared to GPT-2 (cite from Radford et al. (2019)).

G MORE GENERATION EXAMPLES GIVEN LONG PRECEDING CONTEXT

Given the long preceding context exceeding segment length, the contextualized Transformers can generate text condition on the contextual information beyond the segment length, which is different from traditional Transformers.

Preceding Context 1: *Congress is mulling a new law that could force Chinese companies to delist their stocks from American exchanges. Meanwhile, money managers looking for long-term growth opportunities have increasingly turned to China, even before the coronavirus pandemic shocked global growth. More IPOs, investment funds and institutions are set to turn to Hong Kong and mainland China. U.S. pressure on Chinese stocks looks set to accelerate the growth of capital markets in Hong Kong and mainland China, as investors remain intent on chasing opportunities in the world's second-largest economy. The move builds on existing U.S.-China tensions, which began in earnest two years ago on trade, and have since spilled over into technology and finance. In mid-May, U.S. President Donald Trump's administration told a federal pension fund to halt investments in Chinese stocks. Revelation in April of major fraud at Luckin Coffee — which sold itself as a rival to Starbucks in China — accelerated U.S. concerns about lack of transparency into Chinese companies. "Luckin poisoned the well," said Blueshirt managing director Gary Dvorchak, who advises Chinese companies interested in listing in the U.S. "I foresee a very, very difficult environment." At the same time, major international stock and bond index managers have started to include mainland Chinese assets, following years of observation. The inclusion automatically adds some Chinese stocks to many investment funds. Money managers looking for long-term growth opportunities have increasingly turned to China, even before the coronavirus pandemic shocked global growth. Covid-19 emerged late last year in the Chinese city of Wuhan. It has since infected more than 7.3 million people, and killed over 416,000. The outbreak stalled in China by mid-March. Economists expect the country to eke out growth this year, while they predict developed nations like the U.S. will contract. The Asian giant is also home to some of the largest technology start-ups in the world. "To deprive investors from great growth companies is a big mistake, and it will have a meaningful effect on the U.S. financial market," Brendan Ahern, U.S.-based chief investment officer at exchange-traded funds manager KraneShares, said in a phone interview last week. "More investment bankers in Hong Kong, more lawyers, more traders — it will have a true impact on the U.S. finance sector; the New York metropolitan area as a financial hub (for) U.S. capital." The New York Federal Reserve found that the U.S.-China trade war reduced the market capitalization of U.S.-listed companies by \$1.7 trillion, with further hits to investment expected this year, according to a study published in late May. Foreign funds flowing into China In April, allocation to Chinese assets among more than 800 funds held steady from the prior month at almost a quarter of nearly \$2 trillion in assets under management, according to the latest data available from EPFR. The data covers nine categories of stocks listed in mainland China, Hong Kong, Taiwan, the U.S. and Singapore. Chinese government restrictions on cross-border capital flows have made it difficult for foreign funds to access domestic markets, making Hong Kong a more attractive option for international investors wanting to tap China. Underdeveloped regulation on the mainland has also resulted in a rather heavy-handed approach to controlling China's stock markets, which are dominated by retail investors who tend to speculate rather than invest for the long term. For years, many have dubbed the mainland Chinese stock market a "casino." However, analysts say Chinese markets are slowly maturing as more local institutions invest and regulation improves. Foreign investor interest in mainland Chinese stocks has also increased. In late May, the Shenzhen Stock Exchange issued an alert that the foreign investment ratio in three stocks was nearing the 30% limit, the first time such a notice has been issued for three companies, according to China's National Business Daily. Foreign funds accounted for 3.5% of the A shares available for trading, according to data accessed through Wind Information, a financial database. More IPOs in Hong Kong, mainland As U.S. political pressure accelerates, New York-listed Chinese companies are quickly turning to Hong Kong. The U.S. Securities and Exchange Commission is set to hold a roundtable on July 9 to hear views from investors and others on the risks of investing in emerging markets such as China. UBS' in-house regulatory affairs experts expect the U.S. House of Representatives will pass the new legislation against Chinese IPOs by late August, according to a note Tuesday. Ahead of these potential changes, NetEase held a secondary stock offering in Hong Kong on Thursday, while Chinese e-commerce and logistics company JD.com is also planning a secondary listing in coming weeks. "If you're a publicly traded company anywhere in the world this uncertainty is a material risk. There's no way around that," James Early, CEO of investment research firm Stansberry China, said in a phone interview last week. "The second listing idea is going to be a*

lot more palatable than just getting out of the U.S. altogether.” Many Chinese companies have sought the chance to list in New York for the branding benefits, and the opportunity to build capital outside of China’s border controls. Even as Washington-Beijing tensions simmered, Chinese grocery delivery platform Dada went public on the Nasdaq last week. For its part, the Chinese government has wanted to keep its best companies listed closer to home. Last year, the STAR Market launched in Shanghai just months after a directive from Chinese President Xi Jinping. “Shanghai’s STAR market has also lowered the costs of going public in China, removing one of the key reasons why many Chinese companies go public in Hong Kong or abroad,” Jay Ritter, a finance professor at the University of Florida, said in an email. He pointed out that companies in China have faced long delays in getting approval from the China Securities Regulatory Commission. The number of public offerings in China reflects the regulatory changes. In 2018, 105 companies listed on mainland Chinese A share markets, down from 438 the prior year, according to Wind data. When the STAR Market launched last year, it attracted 70 out of the 203 companies that went public, the data showed. For this year through Tuesday, 39 of the 106 public offerings were on the STAR market, according to Wind. The number of public offerings in Hong Kong has also climbed in the last several years, topping 160 last year and 55 for the year so far, according to Wind. The semi-autonomous region has made it easier in recent years for biotechnology companies to list on its exchange. The Hong Kong Stock Exchange is already home to Chinese technology giants such as Tencent and Meituan-Dianping. UBS Securities’ China equity strategy team added in a note that new secondary listings in Hong Kong could still draw investment from U.S. institutional investors that have global operations. Overseas financial firms step up interest International financial institutions have already had their eye on China. The Chinese government has increased efforts to open up the domestic financial industry further to foreign players. The moves are part of a years-long trend, and are also part of the phase one trade agreement signed with the U.S. in January. Critics say Beijing ensured its own financial services industry was well developed before opening the market to foreigners. But many in the business point out that several segments of financial services are still in the early stages of development, such as insurance and asset management. “We’ve seen a lot of interest in the market,” Chantal Grinderslev, partner at Shanghai-based investment management consulting firm Z-Ben, said in an email. “For those clients that are looking into China ... they all realize: In the current global environment, there isn’t an alternative to China.” In one of the latest moves by a foreign firm, Fidelity International – the now independent overseas arm of the U.S. asset management giant – applied last month to set up a wholly owned mutual fund unit. In 2018, the company launched a five-year partnership with Ant Fortune, a subsidiary of Alibaba-affiliate Ant Financial, to study retirement preparedness among people in China. Scully Cui, principal at Bain Greater China, pointed out in a phone interview earlier this week that more and more foreign firms are coming into a Chinese market that is already full of nimble players. “The foreign financial institutions should (move) quickly enough to make the best use of this opening up policy,” she said.

Generated Text 1: "This will strengthen the ability of the Chinese investment managers to invest in China and they need to look at what would be attractive to China in the future through their asset investment investment" to help they develop a more diversified business, the analyst said. It would put the Chinese financial markets in a position to be at the very centre of investor action — an "island" that provides an initial advantage. M.E.T.R. is also planning to invest in Beijing’s emerging markets and the country. "A big role this is going to play in the development of foreign investment in Asia," said M.E.T.R. CEO Yung Hongqing. "When you start to think of China in more general terms, it seems like the investment sector and investing (in Asian investment, especially) is the right market for this. The next part is to figure out how to go forward and the next generation." The Fintech-Growth-E-Commerce business China is the leading growth industry in the world. In June it made a huge increase in it's investment in the world's biggest economic market. The Chinese business world can see more growth in this, and some countries have already started to consider the Chinese as a global destination for business. "It also suggests the potential of Chinese businesses in the market share of China and in many other business areas and it shows (Chinese interest in) the fact that the Chinese industry is a big one." China's current boom for money and services was first started in 2008, in the first year of it in 2014, according to one of the analysts. With a growth in the China business world, "A growing industry is the business growth that is expected to occur in China," said David Hongfeng, associate general partner at the consultancy group China Group. The global economy is "being developed rapidly and it is important to understand the growing demand and the impact in many other parts of China. This is also seen as one of the primary key developments in the realisation of Chinese investments in international markets", according to the investment outlook of China, one

of the five of China's major financial markets, said Hongfeng, a China Business Investment and Research .

Preceding Context 2: *Initial claims are recorded when a person submits a first application for unemployment benefits. Once approved they become continuing claims and are reported with a one-week lag. The report on Thursday is expected to show continuing claims slipped to 20 million in the week ending May 30 from 21.487 million in the prior week. Continuing claims have eased from a record high of 24.912 million in early May. That decline has been attributed to the government's Paycheck Protection Program, part of a historic fiscal package worth nearly \$3 trillion, that offers businesses loans that can be partially forgiven if used for employee salaries. "We are seeing the labor market high on PPP money," said Sung Won Sohn, a business economics professor at Loyola Marymount University in Los Angeles. "Once it runs out we might see a significant increase in layoffs again." Economists caution that neither initial nor continuing claims give an accurate picture. The government has expanded eligibility for unemployment benefits to include gig workers, the self-employed and independent contractors who have been affected by the pandemic. These workers do not qualify for regular state unemployment insurance and must file claims under the Pandemic Unemployment Assistance (PUA) program. Though PUA are filed weekly, they are not included in the initial and continuing claims count. Roughly 36 states are processing PUA claims. Economists recommend focusing on benefit recipients for all programs, which totaled 30 million in the week ending May 16. "The labor market will continue to be under duress as businesses adapt to an economy running well below capacity, resulting in elevated layoffs," said Alex Lin, a U.S. economist at Bank of America Merrill Lynch Global Research in New York.*

Generated Text 2: The state program—the same state that launched PPP, the first public report showing job openings for PPP workers after July 1, 2017. The PPP provides financial assistance to people for those who work part-time. However, the company, which makes more than half of the money it receives this year, does not have an official number through a government data set. The U.S. economic growth is under full employment. While the number in job increases is projected to drop to 6.8 million in 2017, that still holds about \$7.3 billion under the fiscal year's financial and employment data. Economists have reported that the payroll for employment in January and February fell more than three percent. Last week the Bureau of Labor Statistics reported the payroll data show employment in the second month of October at a revised 8.4 percent, the highest quarterly data ever. That has had an impact on the unemployment rates that are already at a record low. In May, the rate had been 3 percent with unemployment at 5.7 percent. That has made an even more significant impact in other industries, the National Public Lumber Employment Index (NOSI), the U.S. labor force survey that ranks jobless in that the NOS was at 7.5 on Friday. NOSI is an overall measure of the number of workers employed in a job and the workers whose employment is in a job, an estimate based on data for most of the most severe recession in the U.S. in at least nine months. The data shows that the U.S. workforce was in the midst of the worst month, which was the first since August 2014. Labor estimates from the NOSI were in March when its peak was only at 6.3 percent, the first month the jobless rate is expected to rise. NOS, which is a labor-specific survey, is a way of looking at jobs of an individual for a year in the labor market with an average weekly job that's an estimate of employment for those at their employment. The NOS is meant to measure actual employment and also to try to make sense of the way people work and in how many are over the job by looking at the jobs job job unemployment for that period is and as of April as it is the number of people job unemployed at those a year or months ago. In April it was at 2.5. In March it is 5.3. Unemployment and employment estimates are at a record for a decade at a high level, a 2 percent increase while that trend for April continues to slow. The U.S. job rate was the lowest at 6.2 in March, well below its lowest level a year ago while the U.S. was at record rate of 7.5 (1.5 in April 2014 and a record low of 6.6 in April 2012) this month. Unemployment has been the third-worst level in a long time, trailing only May 2007, before it began a downturn. The average U.S. unemployment rate for the March to April (an even stronger measure for the month, a 3.5 percentage point difference) is 4.8 percent. The average unemployment in the March to June before the current trend trend downward was 3.8. The jobless rate at all hours in March on the first day of job job-id is 0. The unemployment rate from employment in the first on the week of the month is an average at the month at the first that was a one for May in the January (and the May to the on the December unemployment rate is up at 1.

Preceding Context 3: *The Federal Reserve voted Wednesday to keep benchmark short-term rates near zero. In addition to the rates move, the Fed said it would keep buying bonds, targeting \$80 billion a month in Treasuries and \$40 billion in mortgage-backed securities. On the economy, the Fed sees GDP tumbling 6.5% in 2020 but bouncing back to a 5% gain in 2021. The Federal Reserve kept interest rates near zero and indicated that's where they'll stay as the economy recovers from the coronavirus pandemic. "We're not thinking about raising rates. We're not even thinking about thinking about raising rates," Fed Chairman Jerome Powell said. "What we're thinking about is providing support for the economy. We think this is going to take some time." Along with the rate decision, central bankers projected Wednesday that the economy will shrink 6.5% in 2020, a year that saw an unprecedented halting of business activity in an effort to combat the coronavirus pandemic. However, 2021 is expected to show a 5% gain followed by 3.5% in 2022, both well above the economy's longer-term trend. The central bank repeated its commitment from the April meeting that it "expects to maintain this target range until it is confident that the economy has weathered recent events and is on track to achieve its maximum employment and price stability goals." The Fed also said it will continue to increase its bond holdings, targeting Treasury purchases at \$80 billion a month and mortgage-backed securities at \$40 billion. The Federal Open Market Committee met this week as states begin to reopen and after unemployment saw its worst monthly drop in history followed by its biggest gain. In addition, the meeting comes the same week the National Bureau of Economic Research declared that a recession started in February, ending the longest expansion in U.S. history. Powell said the economic projections were made with the "general expectation of an economic recovery beginning in the second half of this year and lasting over the next couple of years, supported by interest rates that remain at their current level near zero." In early March, the Fed slashed the target range for its overnight funds rate to 0%-0.25%, where it last was during the financial crisis. The rate serves as a benchmark for short-term bank borrowing and also as a guide for most consumer borrowing. Fed officials skipped releasing their quarterly economic projections at March meetings as uncertainty permeated over how long the U.S. would remain in stay-at-home mode and how deep the damage would be.*

Generated Text 3: In a May note to clients, the Federal Reserve said it continues to increase the U.S. interest rate, in a move that would bring to 1-2% and move the rate down, while extending the bond-buying rate that the Fed set and the bank's interest rate. This action would bring to 1 or 2% the U.S. government deficit rate — as indicated earlier, if a more positive market or economic trend was followed by a rebound in the economic recovery. This means the Reserve will continue to raise and then lower this target rate during its next two-month fiscal cycle. The Federal Reserve is tightening monetary policy to create an inflation rate, although there are concerns about the Fed doing enough to ease current-level inflation through the end of 2016 through the end of 2017. The Fed said this year did not see further significant job growth or inflation. "This is a time of sustained inflation," Powell said. Fed, central, and credit markets are expected to see job or earnings growth during the next 5 to 12 months. Investors should consider holding on to the Fed if the U.S. economy continues to grow. The market continued to show strong growth in August, and this season in a positive momentum after the Fed last month reported record losses to their market-adjusted 2.5% rate to 2.3% — the biggest on record. The Fed continued to see job growth on both an economic and labor and inflation-related level. EZG growth, while rising for the first time since February, did no sign of declining. F2 GDP is expected to hit 7.1% and unemployment rate and the first time for a while over a decade. Fed data released Thursday showed interest returns fell below expectations, in a sign of less interest in June, and an upbeat outlook for the market on this year's economic recovery. The long-term inflation rate has fallen to 2 and a rise will last about 1.3% a year, the fourth month of negative interest rate.

Preceding Context 4: *Amazon will temporarily stop providing its facial recognition software to police 729627 CNN Digital Expansion Washington DC 2020, Brian Fung By Brian Fung, CNN Business Updated 6:16 PM ET, Wed June 10, 2020 (CNN Business) Amazon said Wednesday it will stop providing its facial recognition technology to police forces for one year, amid questions about the company's commitment to fighting systemic racism. Amazon said it made the decision in hopes of giving Congress time to "implement appropriate rules" for police use of facial recognition algorithms, which the US government has found are often more likely to misidentify people of color. "We stand ready to help if requested," Amazon said in a blog post. Amazon's Ring subsidiary has partnered with more than 1,300 law enforcement agencies around the country, according to a public tracker on its website. The partnerships have given many agencies far greater visibility into communities through*

the eyes of homeowners and their video surveillance cameras. There appear to be few, if any, limits on police ability to apply Amazon's Rekognition software to video footage obtained from Ring. Last year, in response to CNN's questions on the matter, Ring told CNN that its policies do not govern how police departments handle or store video evidence. The company has previously expressed support for greater regulation of facial recognition. And it recently issued a statement on its website saying "the inequitable and brutal treatment of Black and African Americans is unacceptable." Amazon CEO Jeff Bezos has said on Instagram he is "happy to lose" customers who oppose Black Lives Matter. Amazon's decision follows a letter from IBM to US lawmakers on Monday saying the computing giant will no longer offer facial recognition services and condemning the use of facial recognition "for mass surveillance, racial profiling [or] violations of basic human rights and freedoms." But critics of the company have questioned how Amazon can endorse social justice on the one hand while, on the other, continuing to enable law enforcement agencies with a largely unregulated technology shown to lead to discriminatory outcomes. "It is opportunistic of Amazon to use this moment to make empty and hypocritical statements when it is simultaneously building the backbone for many police departments across the country," Jacinta Gonzalez, field director of the advocacy group Mijente, recently told The Guardian.

Generated Text 4: The problem is the law requires the police to be in a position to use the technology in a way it can, which doesn't match the need. "There has been a pattern of policing when police departments are involved in social media," she added that this "has not happened for some time, although there has been this recent push to make law enforcement police social," while in the past, it has helped law enforcement officials do things like create their own data collection and data data. According to the U.S. Department of Justice, the department can "assess the impact of targeted data, including by searching the social network database." A recent review of other, national data from more than half a decade has shown that police search for users of data on millions of people, and, using the data, identify the law-breaking use to make a police decision based on that data. The department says it can even "citizen-analyze," which allows law enforcement in data surveillance to analyze and analyze information for social and racial context and for how users of that data will impact communities. Police's Data "Cops Data" Crimed The police are used to look at everything from their "real"-time, social media analytics or data collection. The search or seizure data are then used a lot to monitor for crimes, and this comes in the "real" data the use of the police police is doing to look at crime from those databases, and is a good, non-invasive way that makes people more law-abiding. But that data has not been directly collected into this police data. And there can be and the and police data, if it does include some information about our data that does not capture, from what community people are in our neighborhood, our location, to what law enforcement law enforcement would like to say on a law enforcement use of a data system." and in addition, this also "has to say on it's that data, and all people will be a data, that's, and people will use, and and that information and data," and when that becomes in the databases that the police used to the police on, then people will be used for a data analysis.

REFERENCES

- Hao Zhang, Bo Chen, Dandan Guo, and Mingyuan Zhou. WHAI: Weibull hybrid autoencoding inference for deep topic modeling. In *ICLR*, 2018.
- Yulai Cong, Bo Chen, Hongwei Liu, and Mingyuan Zhou. Deep latent Dirichlet allocation with topic-layer-adaptive stochastic gradient Riemannian MCMC. In *ICML*, 2017.
- Jack W. Rae and Ali Razavi. Do transformers need deep long-range memory? In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, ACL 2020, Online, July 5-10, 2020*, pages 7524–7529, 2020.
- Alec Radford, Jeff Wu, Rewon Child, David Luan, Dario Amodei, and Ilya Sutskever. Language models are unsupervised multitask learners. 2019.