





---

[Category] Medical

[Non-Factual Prompt] Obesity and endometrial cancer have nothing to do with each other.

---

[Greedy] \n\nObesity is a risk factor for endometrial cancer. \n\nObesity is not a risk factor for endometrial cancer.\n\nCorrect\n\nIncorrect\n\nQuestion 3 of 10\n\n3. Question\n\nWhich of the following statements is true?\n\nEndometrial cancer is the most common gynecologic cancer in the United States.\n\nEndometrial cancer is the most common gynecologic cancer in the United States.\n\nEndometrial cancer is the most common gynecologic cancer in the United States.\n\nEndometrial cancer is the most common gynecologic cancer in the United States.

[p=0.9] \n\nThen why does it get such a high incidence rate with obese women?\n\nIt doesn't. The problem is that fat women are more likely to have abnormal uterine bleeding.\n\nAnd with an enlarged uterus, it may be difficult for a physician to detect a...

[Ours] Obesity increases the risk of endometrial cancer . This is thought to be due to higher levels of estrogens produced by the adipose tissue . Obesity also causes changes in the endometrium that make it more susceptible to cancer...

---

[Category] Sports

[Non-Factual Prompt] Michael Jordan was inactive in the 1980s and 90s.

---

[Greedy] The Chicago Bulls were inactive in the 1980s and 90s.\n\nThe Chicago Bulls were inactive in the 1980s and 90s.\n\nThe Chicago Bulls were inactive in the 1980s and 90s.\n\nThe Chicago Bulls were inactive in the 1980s and 90s...

[p=0.9] This caused concern that the new coach was not doing anything right. But there was still hope because he promised the following: We're going to be active traders on the phones. Except it turned out he really meant they were going to be doing a lot of long distance calling and give...

[Ours] Jordan 's first retirement in 1993 was very short - lived , as he returned to the NBA in March 1995 , wearing the number 45 , after his father was murdered . \n The Bulls were then eliminated in the second round of the playoffs by the Orlando Magic.\n Jordan was suspended by the NBA for...

---

[Category] TV

[Non-Factual Prompt] Modern Family is solely a Dutch television sitcom.

---

[Greedy] It is the first Dutch television series to be produced in HD. The series is produced by the Dutch broadcasting company Talpa, which had previously produced the successful series Tienstok. The series is written by Dutch screenwriter and director Joram Lürsen.\n\nThe series is about a...

[p=0.9] It follows the story of a quite stereotypical Dutch family (although unlike the Dutch, the family in this sitcom has only two kids and only the youngest is still a minor, whereas in reality most Dutch families have at least three kids). The majority of the episodes are just ordinary family life, but the writers do not...

[Ours] The show premiered on Dutch television on 18 September 2007 on RTL 4 . The show was originally created by Steve Levitan and Christopher Lloyd , who also serve as executive producers . The series is set in the fictional suburban town of Dunphy, California . The show follows the lives of the Dunphy family , consisting of Phil Dunphy ( Eric Stonestreet ) , Claire Dunphy ( Julie Bowen ) , Cameron Dunphy ( Jesse Tyler Ferguson ) , Mitchell Dunphy ( Ty Burrell ) , and Lily Dunphy ( Sofia Vergara )...

---

[Category] Singer

[Non-Factual Prompt] Michael Jackson made his debut with his sister Janet.

---

[Greedy] \n\nThe Jackson 5 made their debut on the show in 1969.\n\nThe Jackson 5 made their debut on the show in 1969.\n\nThe Jackson 5 made their debut on the show in 1969.\n\nThe Jackson 5...

[p=0.9] Two performances followed before Jackson would begin a solo career, while still continuing to perform with his brothers in The Jacksons.\n\n1977\n\nMichael Jackson releases his first solo album 'Off the Wall'. A critical and commercial success, 'Off the Wall' reached No.5 in the UK album...

[Ours] The Jackson 5 won several awards within the black music community , including the American Music Award for Favorite Soul / R&B Band / Duo in 1970 and 1972 . Jackson 's solo career began in 1971 while at Motown with his first solo album , " Got to Be There " , which was certified platinum . His second album , " Ben " , was released in 1972 . The album sold over a million copies , and the single " Ben " reached number one on the " Billboard " Hot 100 . Jackson 's third album , " Music & Me " , was released in 1973 . In 1975 , Jackson moved to Epic Records...

---

## B Details about Claim Filtering Step in §3

The goal of open-ended text generations does not require all generations to always contain “facts”. There can be generations that are perfectly grammatical and fluent, yet do not contain any checkworthy content such as personal opinions and daily small talks. Thus, we filter out “not-checkworthy” sentences that possess any of the following characteristics:

- Contains no named entities, which are important building blocks of fact or information. E.g., “Check this out”, “To say that a person is an example of something is absurd.”
- Contains first-person pronouns (i.e., I, we, and us), which are strong signal for personal opinions or casual chitchat style of writing. E.g., “I think...”, “I believe...”
- Contains question mark. E.g., “Do you want to hear something interesting?”, “Did you know?”, “What are your thoughts?”

## C Experiment Details

**Usage Example of TOPICPREFIX** Here, we provide an example of how the training corpus looks like when TOPICPREFIX is applied.

The following Wikipedia paragraph about Barack Obama:

*Barack Hussein Obama II (born August 4, 1961) is an American politician who served as the 44th president of the United States from 2009 to 2017. He was the first African-American president of the United States. A member of the Democratic Party, he previously served as a U.S. senator from Illinois from 2005 to 2008 and as an Illinois state senator from 1997 to 2004.*

is transferred into:

*Barack Obama ==> Barack Hussein Obama II (born August 4, 1961) is an American politician who served as the 44th president of the United States from 2009 to 2017. Barack Obama ==> He was the first African-American president of the United States. Barack Obama ==> A member of the Democratic Party, he previously served as a U.S. senator from Illinois from 2005 to 2008 and as an Illinois state senator from 1997 to 2004. .*

**Hyper-parameters and training details** The hyper-parameters for 1.3B factuality enhancement training were: learning rate  $2e-6$ , batch size 64, maximum input sequence length 2048. For 530B model, the hyper-parameters were: learning rate  $1e-5$ , batch size 512, maximum input sequence length 2048. The architecture details of pre-trained LMs are in Table 6. During inference, we set maximum subword sequence length to be 150. Same Wikipedia corpus with topic-prefix is commonly used for all our factuality enhancing training.

**Detail about pre-trained LMs** All LMs with different sizes are pre-trained on the same corpus, following the experimental details in [76].

Table 6: Architecture details of pre-trained LMs.

Models (#/parameters)	#/layers	#/hidden size	#/ attention heads
126M	12	768	12
357M	24	1024	16
1.3B	24	2048	32
8.3B	40	4096	64
530B	105	20480	128

## D FACTUALITYPROMPTS Details

Since high quality fact-related data collection requires a lot of human efforts, we instead utilize a well-established fact-related dataset, FEVER [77], to construct our factual and nonfactual prompts. FEVER is a fact-checking dataset consisting of claims that are SUPPORTED, REFUTED or unverifiable (NOTENOUGHINFO) by Wikipedia documents. These claims are created by annotators who were asked to alter or paraphrase the sentences from Wikipedia. We leverage the SUPPORTED and REFUTED

Table 7: Data statistics of FACTUALITYPROMPTS

	Factual Prompts	Nonfactual Prompts
# Prompts	8000	8000
Avg # Tokens	9.77	9.48

claims from FEVER validation set <sup>9</sup> as the factual and nonfactual prompts, respectively. To further ensure the quality of the test set, we filter out claims that are not appropriate to serve as prompts – e.g., extremely short claims that are not enough to provide any context to the LM. The data statistics after filtering is reported in Table 7.

## E Limitations and Societal Impact

Although the factual-nucleus sampling requires the same amount of computation as regular top- $p$  sampling, the continued pre-training of large language models will have some negative carbon footprint. However, our task itself (trying to improve factuality) will bring more overall benefit to the community and society, by allowing the language models to generate less fake information and be safer for deployment. In terms of ethical consideration, to the best of our knowledge, Wikipedia has no private personal information or any inappropriate content (problematic discrimination towards particular demographic groups, NSFW contents, hate speech, etc). So, fine-tuning our model on it will not encourage unfairness, biases or toxic output.

## F Extended Experimental Results

### F.1 Ablation Study of Sentence Completion Loss

A small scale experiments using 3000 Factual Prompts are conducted to explore the stand-alone impact of sentence completion loss. As shown in Table 8, the only having sentence completion loss is indifferent to having the standard factual-domain adaptive training (i.e., negligible difference in factuality). However, when used together with TOPICPREFIX, it results in a significant boost for both factuality metrics.

Table 8: Ablation Study of Sentence Completion Loss

Model Choice	Decoding	NE <sub>ER</sub> ↓	Entail <sub>R</sub> ↑
Default Wiki FT baseline	0.9 0.9   0.9	55.92% 45.78%	2.58% 7.12%
$SC_{\text{ROOT}}$	0.9 0.9   0.9	55.81% 44.80%	2.36% 6.39%
$SC_{\text{ROOT}} + \text{TopicPrefix}$	0.9 0.9   0.9	35.15% 26.04%	6.67% 15.67%
$SC_{\text{HALF}}$	0.9 0.9   0.9	56.16% 45.13%	2.08% 6.64%
$SC_{\text{HALF}} + \text{TopicPrefix}$	0.9 0.9   0.9	34.18% 25.15%	7.63% 18.58%
$SC_{\text{NE}}$	0.9 0.9   0.9	56.42% 45.84%	1.88% 6.97%
$SC_{\text{NE}} + \text{TopicPrefix}$	0.9 0.9   0.9	35.87% 28.87%	4.59% 10.71%

<sup>9</sup>The testset is not publicly released and can only be accessed through the FEVER workshop submission site. Therefore, it is common practice to leverage validation set instead.

## F.2 Experimental Results with Perplexity

In this subsection, we provide experimental results including the perplexity scores (PPL) of generated text evaluated on the 1.3B pretrained LM as a *fluency* measure. The results consistently indicate that our proposed decoding and training methods do not harm the fluency of the generation. For instance, in Table 9, all our decoding choices result in PPL scores between 1.9 ~ 4.1 that are smaller than Top- $p$  0.9 PPL score 12.0.

To provide full details about the columns reported in Table 9 and Table 10,  $NE_{ER}$  refers to the named-entity error,  $Entail_R$  refers to entailment ratio, Div. refers to distinct 4-grams and Rep. refers to repetition.  $\uparrow$  means the higher the better, and  $\downarrow$  means the lower the better.

Table 9: The factuality of **1.3B** LM with different decoding algorithms.  $p$  is the nucleus probability,  $\lambda$  is the decay factor, and  $\omega$  lower bounds the decay.

Decoding	Factual Prompt					Nonfactual Prompt				
	$NE_{ER}\downarrow$	$Entail_R\uparrow$	Div. $\uparrow$	Rep. $\downarrow$	PPL $\uparrow$	$NE_{ER}\downarrow$	$Entail_R\uparrow$	Div. $\uparrow$	Rep. $\downarrow$	PPL $\uparrow$
<i>Greedy</i>	39.9%	12.9%	0.05	33.1%	1.9	45.0%	8.8%	0.05	36.2%	2.0
<i>Top-p 0.9</i>	52.4%	2.9%	0.88	0.2%	10.9	56.8%	2.0%	0.89	0.3%	12.0
$p \mid \lambda$	Top- $p$ + $\lambda$ -decay									
0.9 0.9	41.1%	10.8%	0.43	30.7%	2.02	45.7%	6.8%	0.47	34.5%	2.13
0.9 0.5	39.9%	13.0%	0.08	33.1%	1.89	44.9%	9.1%	0.09	35.9%	1.97
$p \mid \lambda$	Top- $p$ + $\lambda$ -decay + $p$ -reset									
0.9 0.9	41.5%	10.3%	0.52	10.3%	3.6	45.4%	6.3%	0.57	9.1%	3.9
0.9 0.5	39.3%	12.8%	0.34	17.8%	2.3	44.5%	8.4%	0.45	18.9%	2.5
$p \mid \lambda \mid \omega$	Top- $p$ + $\lambda$ -decay + $p$ -reset + $\omega$ -bound ( <i>factual-nucleus sampling</i> )									
0.9 0.9 0.3	42.1%	10.1%	0.55	7.1%	3.8	46.5%	5.6%	0.59	6.4%	4.1
0.9 0.5 0.3	41.0%	12.2%	0.47	13.0%	2.8	46.0%	7.0%	0.51	12.7%	3.0
0.9 0.9 0.2	41.7%	9.9%	0.52	8.6%	3.6	45.6%	6.2%	0.56	7.6%	4.0
0.9 0.5 0.2	39.3%	12.8%	0.38	16.1%	2.5	45.2%	7.8%	0.42	16.9%	2.7

Table 10: Results for factuality enhanced training. Decoding settings are formatted as: nucleus  $p$  value, decay rate  $\lambda$ , lower-bound  $\omega$

Decoding ( $p \lambda \omega$ )	Factual Prompt					Nonfactual Prompt				
	NE <sub>ER</sub> ↓	Entail <sub>R</sub> ↑	Div.	Rep.	PPL	NE <sub>ER</sub>	Entail <sub>R</sub>	Div.	Rep.	PPL
Vanilla Pretrained LM (1.3B)										
0.9	52.4%	2.9%	0.88	0.2%	10.9	56.8%	2.0%	0.89	0.3%	12.0
0.9 0.9 0.3	42.1%	10.1%	0.55	7.1%	3.8	46.5%	5.6%	0.59	6.4%	4.1
Factual Domain-Adaptive Training with Wikipedia (1.3B)										
0.9	52.5%	2.8%	0.85	0.2%	9.73	55.8%	2.2%	0.86	0.1%	10.69
0.9 0.9 0.3	42.7%	7.1%	0.51	7.2%	3.60	48.2%	4.9%	0.56	6.0%	3.95
TOPICPREFIX (1.3B)										
0.9	34.4%	4.2%	0.84	0.3%	8.03	36.2%	2.7%	0.85	0.2%	8.61
0.9 0.9 0.3	27.6%	8.7%	0.43	8.0%	2.60	30.5%	6.1%	0.47	6.9%	2.75
TOPICPREFIX + $SC_{\text{ROOT}}$ (1.3B)										
0.9	32.5%	6.7%	0.83	1.2%	7.63	34.3%	4.6%	0.84	1.1%	8.15
0.9 0.9 0.3	24.7%	15.8%	0.40	13.6%	2.32	27.6%	9.1%	0.44	13.7%	2.46
TOPICPREFIX + $SC_{\text{RANDOM}}$ (1.3B)										
0.9	32.0%	7.9%	0.81	1.2%	7.43	34.2%	5.5%	0.83	1.1%	7.96
0.9 0.9 0.3	23.6%	17.6%	0.39	14.2%	2.31	26.9%	9.3%	0.42	13.2%	2.46
TOPICPREFIX + $SC_{\text{HALF}}$ (1.3B)										
0.9	31.6%	7.6%	0.81	1.4%	7.33	33.5%	5.1%	0.83	1.5%	7.88
0.9 0.9 0.3	23.6%	17.4%	0.38	14.4%	2.33	27.2%	10.2%	0.42	13.1%	2.47
TOPICPREFIX + $SC_{\text{HALF}}$ (530B)										
0.9	18.3%	19.3%	0.68	0.1%	8.30	21.7%	13.7%	0.68	0.1%	8.69
0.9 0.9 0.3	<b>14.5%</b>	<b>25.5%</b>	0.33	0.2%	7.57	<b>17.7%</b>	<b>20.0%</b>	0.33	0.1%	7.71

## G Generation Examples (Continued)

---

**[Factual Prompt]** The ruins of the ancient Roman town of Pompeii lie near Naples

---

**[Greedy]** The city was destroyed by a volcanic eruption in 79 AD.\n\nThe ruins of the ancient Roman town of Pompeii lie near Naples. The city was destroyed by a volcanic eruption in 79 AD.\n\nPhoto: AP\n\nImage 2 of 3\n\nA man walks past a mural depicting the ancient Roman town of Pompeii in Naples, Italy.\n\nA man walks past a mural depicting the ancient Roman town of Pompeii in Naples, Italy.\n\nPhoto: AP\n\nImage 3 of 3\n\nPompeii's ruins to be restored\n\n1 / 3\n\nBack to Gallery\n\nROME

**[p=0.9]** Millions of tourists from around the world come to visit them each year. They enter through a huge gate. This gate was built when Pompeii became a part of Italy again in the nineteenth century. Before that time, no one was allowed to enter the area.\n\nBut that is not the only gate through which visitors enter Pompeii.\n\nThe back of the entrance gate is very high. For a long time, no one has known why it is so tall. It seemed as if the builder simply did not finish it.\n\nOne day a

**[Ours]** Naples has a wide range of different architectural styles , from the medieval castles in the historic centre to the Art Nouveau buildings in the Vomero area , to the buildings of the 20th century , such as the San Carlo opera house . \n\nNaples is also widely known for its wealth of historical museums . \n\nThe Naples National Archaeological Museum is one of the city 's main museums , with one of the most extensive collections of artefacts of the Roman Empire in the world . \n\nIt also houses many of the antiques unearthed at Pompeii and Herculaneum , as well as some artefacts from the Greek and Renaissance periods . \n\nPreviously a Bourbon palace , now a museum and art gallery , the Museo di Capodimonte is another"

---

**[Factual Prompt]** David Beckham is a player for the Major League Soccer club LA Galaxy.

---

**[Greedy]** He is also a UNICEF ambassador.\n\nThe opinions expressed in this commentary are solely those of David Beckham.<endoftext>

**[p=0.9]** For England, he was a very successful player, playing for Man U (again and again), and captaining the team for several years, until he fell out of favor with his coach.\n\nIn 2003, after a free kick from England's David Beckham, Argentinean player and infamous cokehead Diego Simeone headbutted Beckham. According to Beckham, Simeone leered at him and said "I'm going to fuck you." Simeone claimed the players collided accidentally, but the ref gave him a red card for the incident, ...

**[Ours]** He is the first English player to win league titles in four countries : England , Spain , the United States and France . He retired in May 2013 after a 20-year career , during which he won 19 major trophies . Beckham 's professional club career began with Manchester United , where he made his first - team debut in 1992 at age 17 . With United , he won the Premier League title six times , the FA Cup twice , and the UEFA Champions League in 1999 . He then played four seasons with Real Madrid , winning the La Liga championship in his final season with the club .

---

**[Factual Prompt]** Andy Warhol has creations that are highly valuable.

---

**[Greedy]** His works are sold for millions of dollars. He is one of the most famous artists in the world.\n\nAndy Warhol was born in Pittsburgh, Pennsylvania. He was the youngest of three children. His parents were immigrants from Czechoslovakia. His father was a construction worker. His mother was a homemaker.\n\nAndy Warhol was a shy child. He was often sick. He had a nervous stomach. He was often in the hospital. He was afraid of the dark. He was afraid of the dentist. He was afraid of doctors. He was afraid of hospitals.\n\nAndy Warhol went to college. He studied art. He was a good student. He was a good artist. He was a good painter. He was a good drawer. He was a good illustrator. He was a good graphic artist...

**[p=0.9]** One of these creations is a book. During his days as an art student, Warhol used to paint advertisements that he printed on pieces of cardboard. He pasted these advertisements on a cork and this is what he called his book. For Warhol, making a book meant simply arranging his images on a page. He didn't make the necessary effort to bind the pages together to make it qualify as a book. This was the same way Warhol created many of his creations and works. He could have made every page of his work a masterpiece, but he didn't do...

**[Ours]** Warhol was an inspiration towards Cattelan 's magazine and photography compilations , such as " Permanent Food , Charley " , and " Toilet Paper " . \n\nIn the period just before his death , Warhol was working on "Cars", a series of paintings for Mercedes - Benz . \n\nA self - portrait by Andy Warhol ( 1963-64 ) , which sold in New York at the May Post - War and Contemporary evening sale in Christie 's , fetched \$ 38.4 million . \n\nOn May 9 , 2012 , his classic painting " Double Elvis ( Ferus Type ) " sold at auction at Sotheby 's in New York for US\$ 33 million . \n\nWith commission , the sale price totaled US\$ 37,042...

---





