Counterfactual Memorization in Neural Language Models

Chiyan Zhang
Google Research
chiyuan@google.com

Daphne Ippolito
Carnegie Mellon University
daphneicmu.edu

Katherine Lee
Google DeepMind
katherinelee@google.com

Matthew Jagielski
Google DeepMind
jagielski@google.com

Florian Tramèr
ETH Zürich
florian.tramer@inf.ethz.ch

Nicholas Carlini
Google DeepMind
ncarlini@google.com

Abstract

Modern neural language models that are widely used in various NLP tasks risk memorizing sensitive information from their training data. Understanding this memorization is important in real world applications and also from a learning-theoretical perspective. An open question in previous studies of language model memorization is how to filter out “common” memorization. In fact, most memorization criteria strongly correlate with the number of occurrences in the training set, capturing memorized familiar phrases, public knowledge, templated texts, or other repeated data. We formulate a notion of counterfactual memorization which characterizes how a model’s predictions change if a particular document is omitted during training. We identify and study counterfactually-memorized training examples in standard text datasets. We estimate the influence of each memorized training example on the validation set and on generated texts, showing how this can provide direct evidence of the source of memorization at test time.

1 Introduction

Modern neural language models (LMs) have achieved impressive results in generating high quality text [e.g. Brown et al., 2020, Zhang et al., 2022, Chowdhery et al., 2022, OpenAI, 2023] and have led to breakthroughs in many downstream natural language processing tasks [Devlin et al., 2019, Raffel et al., 2020b, Bommasani et al., 2021]. The paradigm of taking a single large-scale pre-trained model and fine-tuning it for many tasks motivates the study of these models’ ability to generalize by avoiding memorizing their training data. Moreover, memorization of sensitive user information or copyrighted materials in the training data [Carlini et al., 2020, Vyas et al., 2023, Lee et al., 2023] leads to practical concerns in real world applications.

Previous work on memorization in neural language models demonstrated the ability to extract memorized training data, including sensitive data such as phone numbers and usernames [Carlini et al., 2020, Ziegler, 2021, Carlini et al., 2019, Henderson et al., 2017, Thakkar et al., 2020, Thomas et al., 2020]. One issue with these extraction attacks is that they primarily identify “common” and frequently occurring strings in the training set. For example, as shown in the analysis of Lee et al. [2021], near-duplicate training examples, which are very common in standard text corpora, account for a large majority of the memorized content. To filter out such commonly occurring strings from all memorized texts, previous work applied various heuristic rules to distinguish frequently-occurring sequences from memorization of isolated pieces of information.

In this paper, we propose a principled causal perspective to disentangle memorization of common vs rare data, by directly tying a model’s predictions to the presence or absence of individual training examples. We define counterfactual memorization as a measure of the change in a model’s prediction when a particular example is excluded from the training set. Counterfactual memorization accounts for the commonality of an example as removing one instance of a text that is common across multiple documents will have a minor effect on the model’s prediction on that text. The mathematical formulation of counterfactual memorization extends a prior definition of label memorization in classification models [Feldman, 2020] to the context of neural language modeling.

Formally, a training document $x$ is considered counterfactually memorized, when the language model predicts $x$ accurately if and only if the model was trained on $x$. This allows us to construct a procedure to quantitatively measure the memorization of isolated pieces of text, whose sole presence in the training dataset have a large effect on the model’s predictions.

Following Feldman and Zhang [2020], we further extend this definition to counterfactual influence, which measures the influence of a memorized training text sample on another text example. Counterfactual influence allows us to trace the source of information for a model’s predictions, by locating the training example(s) which significantly contributed to it. With these tools, we study memorization across several standard text datasets. Our main contributions are as follows:

1. We define counterfactual memorization in neural LMs which gives us a principled perspective to distinguish memorization of “rare” and “common” information in neural LMs (Section 3).
2. We estimate counterfactual memorization on several standard text datasets, and confirm that rare memorized examples exist in all of them. We study common patterns across memorized text and the memorization profiles of individual internet domains. (Section 4).
3. We identify an inverse correlation between number of duplicates and counterfactual memorization as compared with previous definitions of memorization (Section 5).
4. We extend the definition of counterfactual memorization to counterfactual influence, and study the impact of memorized examples on the test-time prediction of the validation set examples and generated examples (Section 6).

2 Related Work

Previous work analyzed the memorization of large language models on sensitive information (e.g. phone numbers) in the training data [Carlini et al., 2020, Ziegler, 2021] or synthetically injected “canaries” [Carlini et al., 2019, Henderson et al., 2017, Thakkar et al., 2020, Thomas et al., 2020]. However, not all the memorized texts are equally interesting — as confirmed in a later study [Lee et al., 2021], near-duplicated training examples are very common in standard text corpus, and those commonly occurring phrases contribute significantly to memorized texts. In order to distinguish “common” memorization of common phrases or public knowledge from “rare” memorization of private, rare information, various heuristics were adopted in previous investigations. Our paper proposed a principled perspective towards this problem. Our intuition comes from psychologies studies that categorize human (declarative) memory into episodic memory [Tulving, 1983] of specific contents of individual events, and semantic memory [Squire, 1992] about general knowledge like grammars and factual information. We would like the models to obtain semantic memory but avoid episodic memory. The capture the latter, we proposed a notion of counterfactual memorization. The mathematical formulation of counterfactual memorization is borrowed from a notion of label memorization in Feldman [2020] and adapted to the context of neural LMs in this paper. This formulation has been studied empirically in the context of computer vision in Feldman and Zhang [2020]. In a follow up work, Ilyas et al. [2022] showed that it is possible to fit a datamodel to predict the outcome of training a model on a specific training subset and evaluating on a specific input. However, this procedure requires training a massive number of models (e.g. 300,000 for CIFAR-10) on random subsets of the training data, thus is computationally infeasible for the scale of language models considered here.

The general idea of measuring model behavior on held-out training data is common in machine learning. In cross validation, held-out data is used to estimate the test performance for model selection; in learning theory, leave-one-out stability was shown to be deeply connected to generalization [e.g. Mukherjee et al., 2006]; in differential privacy, the worst case performance difference of models trained on two “neighboring” datasets (identical except a single example being held-out or replaced)
quantifies the privacy guarantee of a learning algorithm [Dwork et al., 2014, Nasr et al., 2021, Jagielski et al., 2020]. Most previous work aimed for an overall measurement, while our paper focused on characterizing the behaviors of individual examples.

We estimated a counterfactual influence to study how a memorized training example impact the model prediction at test time. Influence functions have been used in statistics to assess robust estimators since Hampel [1974]. Previous papers adopted it to analyze neural network predictions [Koh and Liang, 2017, Koh et al., 2019]. However, the estimation was found to be computational expensive and fragile [Basu et al., 2021]. Pruthi et al. [2020] tracks the gradient updates during training to estimate the influence from a training example; Feldman [2020], Feldman and Zhang [2020] use aggregated statistics from multiple models independently trained on heldout data subsets to estimate the influence. Further extensions were shown to work well on detecting mislabeled data in classification problems [Wang and Jia, 2022] and characterizing hallucinations in Neural Machine Translation [Raunak et al., 2021]. Alternative methods also looked at simple data statistics (e.g. co-occurrence counts) without model re-training to infer the causal effects on language models’ predictions [Elazar et al., 2022]. In this paper, we adapt the approach from Feldman [2020], and formulate counterfactual influence directly with subset sampling, as oppose to leave-one-out influence. We also extend the estimation to assess the influence on generated examples.

Counterfactual is an important notion in statistical causality [Pearl et al., 2000, Rubin, 2005, Pearl, 2009, Imbens and Rubin, 2015] useful for studying causal probabilistic inference under alternative conditions. Such counterfactuals may or may not be directly testable (e.g. a counterfactual treatment in medical studies). In this paper, we directly measure the counterfactual influence of a training example by comparing the behavior of the model trained with and without that example.

### 3 Counterfactual Memorization

To quantify memorization of rare details of a specific training document, we define the following notion of counterfactual memorization. The mathematical formulation is borrowed from Feldman [2020], where it was originally proposed to quantify label memorization in multi-class classification problems. We extend it to the context of unsupervised neural language modeling.

**Definition 3.1 (Counterfactual Memorization).** Given a training algorithm \(A\) that maps a training dataset \(D\) to a trained model \(f\), and a measure \(M(f, x)\) of the performance of \(f\) on a specific example \(x\), the counterfactual memorization of a training example \(x\) in \(D\) is given by

\[
\text{mem}(x) \triangleq \frac{\mathbb{E}_{S \subseteq D, x \in S}[M(A(S), x)] - \mathbb{E}_{S \subseteq D, x \not\in S}[M(A(S), x)]}{\text{mean performance on } x \text{ when trained with } x - \text{mean performance on } x \text{ when not trained with } x},
\]

where \(S\) and \(S'\) are subsets of training examples sampled from \(D\). The expectation is taken with respect to the random sampling of \(S\) and \(S'\), as well as the randomness in the training algorithm \(A\).

That is, our memorization definition compares the difference between two expected performance measures on a given example \(x\). On one side, we compute the expected performance of a model when trained on datasets that contain the example \(x\), and, on the other side, we compute the expected performance of a model when trained on datasets that do not contain the example \(x\). Throughout this paper we use per-token accuracy as the measure \(M\). In other words, we ask the model to predict the next token based on the groundtruth context (preceding tokens), measure the 0-1 loss of the argmax token prediction, and then average it across all predicted tokens.

The expectations in Equation (1) can be empirically estimated via sampling. Specifically, we train \(m\) different models on independently sampled subsets \(S_1, \ldots, S_m\) of equal size \(|S_i| = r|D|\) for a fixed \(r \in (0, 1)\). We then divide these models into two groups: the first group contains all models trained on subsets \(S\) where \(x \in S\); and the second group are all models trained on subsets \(S\) where \(x \not\in S\). We take the average performance on \(x\) in the two groups separately and compute the difference between the two:

\[
\text{mem}(x) \triangleq \text{mean}_{i : x \in S_i}[M(A(S_i), x)] - \text{mean}_{i : x \not\in S_i}[M(A(S_i), x)].
\]

This difference quantifies how the presence or absence of the example \(x\) in a model’s training set affect the model’s performance on \(x\). If there is a large difference between including an example in the training set versus not including it, then we consider this example counterfactually memorized.
For each $x$, we refer to models trained with $x$ in the training set ($\{ A(S_i) : x \in S_i \}$) as IN models and the models $x$ was not trained on ($\{ A(S_i) : x \notin S_i \}$) as OUT models. Note we do not need to retrain a model for each example $x$. Instead, we train $m$ models once on random subsets of $D$, and compute the estimation (Equation 2) for all examples using the same set of $m$ models. Ilyas et al. [2022] recently showed that it may also be possible to directly predict these scores using a regression model, yet this approach is computationally prohibitive for large language models.

4 Analyzing Counterfactual Memorization

We estimate and analyze counterfactual memorization of training examples in three standard text datasets: RealNews [Zellers et al., 2019], C4 [Raffel et al., 2020a] and Wiki40B:en [Guo et al., 2020]. Unless otherwise specified, we use Transformer-based language models [Vaswani et al., 2017] equivalent to (decoder only) T5-base [Raffel et al., 2020b] with $\sim112$M parameters. To save computation and enable more direct comparisons across datasets, we truncate the training set for each datasets by taking the first $2^{21}$ documents. To estimate counterfactual memorization, we train 400 models for each dataset, each on a random $25\%$ subset of the training examples. In practice, we use a hash-based filtering mechanism to efficiently approximate random subset sampling (details in Appendix G), as the data loading APIs for large text corpora generally support only sequential visits to examples with limited shuffling and subsampling capability within a window.

We train each model for 60 epochs using the Adam optimizer [Kingma and Ba, 2015] with learning rate 0.1 and weight decay $10^{-5}$. For C4/RealNews/Wiki40B:en, respectively, our models converge to an average per-token accuracy of 44.21%/47.59%/66.35% on the subsampled training set, and 27.90%/31.09%/49.55% on the validation set. On average, the models start to overfit at around epoch 5, as indicated by the signal that the validation accuracy starting to decrease.

4.1 Distribution of Memorization

Table 1 shows examples from the RealNews training set sampled at various memorization levels. Examples with the highest memorization are generally unconventional text such as all-capital letters, structured formats (i.e., tables or bullet list), and multilingual texts. After those artificial examples, examples with intermediate-to-high memorization are most often news reports of specific events. One of our main goals is to be able to separate memorization of such examples containing details of specific events from memorization of common facts or highly duplicated template texts. Indeed, templated documents with many near-duplicate copies in the training data generally have low counterfactual memorization. C4 and Wiki40B:en have similar trends. Interestingly, though Wikipedia articles are less likely to be auto-generated from templates than the web in general, we do observe repetitive patterns in low-scoring documents, such as "_START_ARTICLE_<place name>, Virginia _START_PARAGRAPH_<place name> is an unincorporated community in <county name>, in the U.S. state of Virginia."

To visualize the distribution of memorization, we plot 2D histograms in Figure 1, where the x-axis shows the difference of IN-accuracy and OUT-accuracy (i.e. the counterfactual memorization), and the y-axis shows the sum of the two, which we term “simplicity”. A simple example is one that is scored highly regardless of whether a model saw it during training. The histograms are plotted in log scale to better visualize the exponential decay in the tail for high memorization and simplicity levels.

From the 2D density plots, we find that easy examples tend to have low memorization. However, there is no simple linear correlation. Peak memorization occurs for examples of intermediate simplicity. For the hardest examples, the memorization scores are low, because even the IN-models could not learn them well. Many hard examples consist of ill formatted text or contained foreign languages. As a result, in Wiki40B:en, which contains higher quality texts, the lower bound of the histogram is higher than the other two datasets (Figure 1). Interestingly, the choice of data has a relatively minor effect on memorization: the shape of the memorization histogram is generally consistent across the three datasets; the range of memorization values is only slightly compressed for Wiki40B:en.

\[1\] Modern language models are usually trained for fewer epochs if the training set is massive. Since we have a smaller subsampled training set, we train the models for more epochs to allow the models to fit the training data sufficiently to study memorization effects.
Table 1: Examples of RealNews training set sampled at high, intermediate and low memorization. The URL of each document is included at the beginning of each example. [33] indicate omitted text for brevity. In the last block, two near-duplicate examples are shown; the yellow highlights in the last block indicate differences.

```
<table>
<thead>
<tr>
<th>Index</th>
<th>mem</th>
<th>Text</th>
</tr>
</thead>
</table>
| 2008855 | 0.6540 | [33] ← THE AMERICAN JEWISH CONGRESS ANNOUNCED TODAY THE PUBLICATION OF A REPORT ON JEWISH NON-EMPLOYMENT AS A RESULT OF ECONOMIC DISCRIMINATION. [33] THEREAFTER ONE OF THE DEPARTMENTS OF A.T. & T. "ALMOST UNPRECEDENTEDLY" ENGAGED A JEWISH APPLICANT.
| 2085736 | 0.7755 | [33] ← RECIPE. Chinese Pork & Vegetable Soup with Wonton Noodles Chinese Pork & Vegetable Soup with Wonton Noodles (about 1-1 1/4 pound size), cooked and cut into 1/2-inch cubes* 5 cups low-sodium chicken broth 1 cup water** 1/4 cup [33] Makes 6 servings (about 1 1/2 cups each) Recipe by PorkBeInspired.com with adaptations by culinary dietitian & nutritionist Kim Galeaz. RDN CD
| 1680600 | 0.5807 | [33] ← Language English [33] acknowledgement of country [33] ←[33] Arabic text [33] I would like to acknowledge that this meeting is being held on the traditional lands of the (appropriate group) people, and pay my respect to elders both past and present. [33]
| 2074085 | 0.2835 | [33] ← A Texas honors student punished for saying that homosexuality was wrong has had his suspension rescinded [33] Western Hills High made the correct decision in reversing their course of action. "The decision to rescind the suspension is the correct one. The suspension was wrong and improper," said Staver. "I applaud the student for standing up. We stood with him to resist an unjust suspension and we are pleased that suspension has been reversed." [33] Liberty Counsel will continue the right to exercise freedom of conscience and religion," said Staver. “These instances are increasing and will continue to increase unless Christians and people who love liberty stand up and resist this intolerance.”
```

![Figure 1](image-url)

**Figure 1:** The joint distribution of counterfactual memorization (X axis) and simplicity (Y axis), where simplicity is measured as the overall accuracy for an example across all models. (Histograms are in log-scale).

To investigate these effects, we visualize the 95th percentile memorization score in each web domain against the number of examples in that domain for RealNews (Figure 2a) and C4 (Figure 2b). C4 contains many more domain names than RealNews since the latter is collected only from news websites. For both datasets, the domains with a large number of crawled documents show a smaller variance in the 95-percentile values, while “smaller” domains depict a wide range of variety in memorization profiles. The memorization profiles of a few representative domains are visualized in Figures 2c and 2d. The domains we selected for visualization are: the largest domain (blue), the domain with highest 95 percentile memorization (orange), and two domains that have more than 1000 and 50 articles in RealNews and C4 respectively (green and red).

In RealNews (Figure 2c), reuters.com contains the largest number of documents but low memorization scores on average. The domain digitallibrary.un.org, the United Nations Digital Library, has high memorization scores potentially because it contains many multilingual documents. We have observed that less frequently occurring tokens, like those in foreign languages or ALL-CAPITAL words tend to cause high memorization. Similarly, flattened structured data (e.g. tabular texts) also deviates significantly from normal English texts and potentially leads to high memorization, as demonstrated by zap2it.com, a website for TV program listings. On the other hand, hotair.com is a
news commentary website that frequently quotes other major news articles. This may lead to duplicate text in the dataset which we suspect contributes to its overall lower memorization distribution.

The observations are similar on C4: blogspot.com contains a large number of documents in the training set with only moderate amounts of memorization; zh.wikipedia.org and buckinghamautos.com.au have high memorization due to foreign (Chinese) or structured (car sales listings) text; and www.unitedstateszipcodes.org has very low memorization scores because common templates are re-used to generate similar pages for individual zip codes.

4.2 Number of Models Needed

To evaluate the impact of a single training example, one may wish to train two models that differ only in that single example. In practice, the stochasticity in a single run of common training algorithms (e.g. SGD) produces too low signal-to-noise ratios to be useful for such estimation. Moreover, leave-one-out estimation means a separate pair of models needs to be trained for each training example, which is computationally costly. Therefore, we formulated our estimation in Section 3 by accumulating statistics from $m$ models independently trained on random training subsets. In our experiments, we set $m = 400$. To understand how sensitive our results are to $m$, we analyze the rankings produced by distinct sets of models of size $m$. We vary $m$ from 6 to 192, and partition our set of 400 models into up to 10 sets of $m$ models (e.g. for $m = 192$, we construct 2 partitions, and for $m = 6$, we construct 10). We then compute the Spearman’s R between these partitions to measure the agreement between the rankings produced by each partition. If the rankings are very similar (have Spearman’s R close to 1), then this number of models is reliably estimating the true ranking of memorization scores. We plot these Spearman’s R values in Figure 3a. Even at 96 models, this correlation begins to plateau near 1, lending confidence that 400 models is sufficient for reliable estimation of memorization scores. See Appendix D for more analysis on the sensitivity to $m$.

4.3 Impact of Number of Training Epochs

As expected, the overall amount of memorization grows consistently with the number of epochs of training (Figure 3b). This makes sense since training for more epochs increases overfitting. As training progresses, we also see an increasingly long tail of examples with high memorization scores. On RealNews, about 59% of examples had consistently increasing memorization scores across all epochs considered. There were no examples whose memorization decreased in a significant way over training (all observed decreases can be attributed either to noise or to instability early in training). Only 0.5% of examples stayed completely un-memorized with scores which never rose above 0.1,
while 85% of examples had memorization scores which never rose above 0.2. Figure 3c shows the fraction of memorized examples as training progresses, at several thresholds of memorization. We can see that more training epochs significantly increases memorization.

5 Duplicate Text and Memorization

One of the goals of evaluating counterfactual memorization is to identify examples that have a low number of duplicates yet whose presence versus absence in the training data has a large effect on the model. Here, we perform a quantitative study of the (anti-)correlation between duplication and counterfactual memorization compared with the positive correlation between duplication and the “generation-time memorization” definitions of memorization used by Lee et al. [2021], Carlini et al. [2022], Kandpal et al. [2022].

Following the method from [Lee et al., 2021], we first use MinHash [Broder, 1997] to identify near-duplicate examples in RealNews train set. We consider a example a duplicate if it has an normalized edit similarity of greater than 0.7 (definition included in Appendix I). Out of 2.01 million examples, ~38,000 were identified as being a near-duplicate with at least one other example. Among these frequently-occurring examples, the Pearson correlation between an example’s counterfactual memorization score and the number of near-duplicates for that example is -0.39; in other words, memorization does quantitatively decrease when data is repeated more often.

In Figure 3d we can see that examples with a large number of near-duplicates have smaller memorization scores. Counterfactual memorization primarily differentiates amongst examples with a few number of duplicates. This makes sense given that examples with lots of near duplicates would likely have their near duplicates in OUT-models. This is to be contrasted with “generation-time memorization” (discussed in Section A) that measures the textual overlap between model generated texts and the training documents. There, the number of occurrences strongly correlate with the measured memorization [Carlini et al., 2020, Lee et al., 2021, Kandpal et al., 2022]. Counterfactual memorization measures a fundamentally different type of memorization from simple textual matching considered in prior work, providing information about how easy or hard a training example is in the context of the rest of the training set. In Table 1 we can see this effect qualitatively: sequences with near-duplicates in the training set tend to have low counterfactual memorization (as expected).

6 From Memorization to Influence

Counterfactual memorization identifies training examples that contain rare information not conveyed by other examples. A natural question to ask is whether a model would leak the information in a memorized example during inference. Previous paper studies membership inference attack [Shokri et al., 2017, Sablayrolles et al., 2019, Long et al., 2020] where an attacker tries to figure out if a particular example exists in the training set. In this paper, we consider standard model evaluation without adversarial attackers, and quantify “does seeing a particular training example strongly influence the prediction on a validation example?” Another way of asking this is if a single example in the training set has an large and over-representative impact on the prediction of a validation
An empirical estimation of the influence can be computed similarly to counterfactual memorization where $S$ is a subset of training examples sampled from $D$. The expectation is taken with respect to the random sampling of $S$, as well as the randomness in the training algorithm $A$. Here $x'$ can be an example from the validation set or test set, a generated example or a training example.

An empirical estimation of the influence can be computed similarly to counterfactual memorization by uniformly sampling $m$ subsets $S_1, \ldots, S_m$ from $D$, where $|S_i| = r|D|$, and calculating

$$\text{infl}(x \Rightarrow x') \triangleq \frac{1}{m} \sum_{i=1}^{m} \left[ \frac{\text{mean}[M(A(S_i), x')]}{\text{mean}[M(A(S_i), x')]} \right].$$

This measures how much a training sample $x$’s presence influences the prediction of a different example $x'$. Note, $\text{mem}(x) = \text{infl}(x \Rightarrow x)$, i.e., counterfactual memorization is self influence.

**Influence on Examples of the Validation Set.** With the same models trained for estimating memorization, we can estimate the counterfactual influence on the validation set according to Equation (4). For each example in the validation set, we can estimate the influence on it from each training example. Figure 4a shows the distribution of influence from all training example on three different examples from the validation set. The green example was randomly chosen and represents the behavior for most validation examples: it receive close-to-zero influence from all the (individual) training examples. The blue and orange examples were sampled to have high and intermediate maximum influence. Each of them has one (or a few) strong influencer from the training set, as indicated by the bars to the right of the histogram. They also only receive tiny influence from all the rest of the training examples, though the variance of influence is larger than for the green example.

Intuitively, most training examples will have small influence on validation set examples because the models learn distributional patterns shared across many training examples, and *individual* training examples tend to have insignificant influence here. However, a training example $x$ with high counterfactual memorization contains rare information that are not shared with other examples. Therefore, if a validation set example $x'$ contains similar information, $\text{infl}(x \Rightarrow x')$ could be large. Figure 4b shows the relationship between memorization and influence by plotting $\text{mem}(x)$ of each training example $x$ against its maximum influence $\text{max}_{x'} \text{infl}(x \Rightarrow x')$ on $x'$ across the validation set.

Consistent with our intuition, examples with small memorization scores have small max-influence scores. Larger influence scores on the validation set generally requires larger memorization scores.
of the training example itself. However, not all training examples with large memorization scores lead to large influence scores. In particular, the max-influences drop significantly for examples with memorization larger than 0.4. One potential reason is that many examples with very high memorization are simply low quality text, so memorization is required in order to learn them, but they do not encode anything interesting that could influence a validation example. On the other hand, even if a memorized example encodes some rare and useful information, the max-influence could still be low because the validation set does not contain a relevant document. This is especially true given that all datasets have considerably smaller validation sets than training sets.

Table 2 shows train-validation example pairs from RealNews sampled at different influence value ranges. We found that the train-validation pairs with the highest influence are almost identical, except some superficial differences, such as different handling of quotation / em dash marks. As we move to intermediate influence ranges, we commonly found reports on the same events. Large paragraphs of identical text indicate that one document might be citing the other or both citing from a third party. At low influence, two types of correlations are commonly observed: 1) templated texts with high similarity—the reason for a low influence is that there are many similar training examples that split the influence; 2) superficially related documents due to a shared prefix such as a shared substring of some common knowledge.

Influence on Generated Texts. The influence estimation is not restricted to the validation set. We can also estimate influence on generated examples. In this section, we evaluate on the publicly released generations from the Grover models [Zellers et al., 2019] trained on RealNews. Specifically, we take the generations from Grover-Mega (p=0.96), a 1.5-parameter-beam model trained on the RealNews
dataset. Comparing with the train-validation influence in Figure 4b, the histogram (c.f. Figure 10 in Appendix.) decays faster as max-influence grows. Moreover, the value range of max-influence is also twice smaller. The reason that we did not find a lot of highly influenced generated examples are two fold: 1) there are only 24,576 generation in the public release, which is much fewer than the validation examples. As a result, the corresponding example of many memorized training examples do not get sampled in the generations. For comparison, previous work [Carlini et al., 2020, Lee et al., 2021] generated 100,000+ examples to identify memorization in generation. These approaches also count duplicates in the training set, which counterfactual memorization filters out. 2) The Grover model was trained on the full RealNews training set, while we have restricted our analysis to the first 2M training examples. There could be potentially more high influence training examples that are missed in our calculation.

7 Summary and Discussion

We studied memorization in neural language models. We formulated a notion of counterfactual memorization as a tool that can systematically ignore “common” memorization such as general knowledge (e.g. “Paris is a city in France”) and captures memorization of rare, specific information (e.g. description of a specific episode of event) present in the training examples. We conducted experiments on three commonly used text corpus in language modeling and found memorization in all of them. We further analyze the per-domain memorization profiles for Internet-crawled data, and found that different sources could have substantially different memorization profiles.

Furthermore, we analyzed how memorized training examples could impact the model predictions at test time via counterfactual influence. We found that for examples from both the validation set and the model generated texts, the model predictions could be drastically different depending on the presence or absence of a particular training example with high memorization.

Limitations. This study mainly focus on English datasets. While we expect the characterization of memorization would be similar when evaluated on corpus of other (natural) languages, new patterns might be observed on multilingual data or more structured domains such as programming languages.

Both the neural language models and training sets used in this work are orders of magnitude smaller than modern standards such as GPT-3 [Brown et al., 2020], GPT-4 [OpenAI, 2023] and PaLM-2 [Google, 2023]. Moreover, we only conducted preliminary investigation of the dynamics of counterfactual memorization during training. Although our experiments effectively estimated and detected memorization, we suspect more interesting examples might emerge if larger, more capable models are analyzed. For example, currently when the information from a memorized training example is leaked in the prediction of a strongly influenced test example, it can usually be explained by a high text overlap between the training and test examples. For models with deeper understanding of languages, we suspect that strong influence could be observed even between documents that have no direct text overlap but that encode similar semantic information.

In order to test this, it will be necessary to scale our framework to larger models and datasets. Moreover, it will be necessary to construct datasets where semantically similar but textually different document pairs exist. One potential source to construct such datasets would be versioned Wikipedia articles—two versions of the same article with large time span or edit distance may contain semantically similar (but paraphrased) information. Such a dataset of paraphrased text pairs would be more broadly useful to understand the ability of different models to disentangle text content and form—by measuring the influence of one piece of text on a paraphrased piece of text.

Counterfactual memorization enables us to identify examples that whose presence or absence has a large impact on the model and the model’s ability to score and generate other text. The privacy risk for this is low since in order to perform this analysis, one would need to already have access to the dataset and the ability to train models.

Acknowledgments. The authors would like to thank Samy Bengio, Christopher A. Choquette-Choo, Ethan Dyer, Michael C. Mozer, Behnam Neyshabur, Andrew Nystrom, and Hanie Sedghi for constructive discussions and feedback. The authors would like to thank Andrew Nystrom for assistance with MinHash-based near-duplicate detection.
References


Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramer. Membership inference attacks from first principles, 2021.


A Difference Between Counterfactual and Generation-Time Memorization

Many definitions of memorization operate at *generation-time*: a sequence of generated text is marked as memorized if a sufficient amount of overlap is found in the training dataset [Carlini et al., 2020]. When the training data is not available, heuristic-based methods comparing language model perplexities are used to predict whether a generation contains memorized content [Carlini et al., 2019, Thakkar et al., 2020, Thomas et al., 2020, Carlini et al., 2020, Zanella-Béguelin et al., 2020]. One difficulty with these approaches is that generation-time instances of memorization are strongly correlated with the number of similar or near-duplicate examples in the training set. As observed in Lee et al. [2021], large clusters of near-duplicated examples do exist in common language datasets, dominating memorization detected in generated text. Generation-time methods for measuring memorization are forced to design heuristics to avoid simply identifying these uninteresting instances of memorization.

In contrast, the counterfactual memorization we study in this paper handles the issue of near-duplicates automatically without the need for heuristics. For a training example, \( x \), with many near-duplicate copies in the training set, \( \text{mem}(x) \) will be small (because other samples \( x' \approx x \) will be present in the training dataset whether or not \( x \) is). This does not mean that counterfactual memorization is the opposite of generation-time memorization. An example, \( x \), with high \( \text{mem}(x) \) may have a high chance of being generated if a model is appropriately prompted, despite and possibly because it is rare, and thus the example is considered memorized by both definitions. In summary, generation-time memorization measures the chance a model will directly copy from training examples, while counterfactual memorization aims to discover rare information that is memorized.

B Average Accuracy of IN models vs OUT models

Figure 5 compares the per-token accuracy between the IN models and OUT models for the training examples from three different datasets. Counterfactual memorization is estimated by taking the difference between the average IN-accuracy and the average OUT-accuracy. Thus, the examples closer to the upper left corner are more counterfactually memorized, while the examples near the diagonal are not.

C The Impact of Data Deduplication on Memorization

To investigate the impact of data deduplication on counterfactual memorization, we compared C4 with C4-NEAR Dup [Lee et al., 2021], which is derived from C4 with deduplication using approximate document matching. Figure 7 compares the distribution of memorization between the original C4 and the deuplicated dataset. We did not find significant difference between the two datasets. One potential reason is that the deduplication criterion was relatively conservative, which removed only \( \sim 3\% \) of the training examples. In fact, we can still easily see near duplicate examples in C4-NEAR Dup among examples with low memorization, as shown below:

**Example 1380925** (mem = 0.0374) link ▷ This is a placeholder page for Joshua Baldridge, which means this person is not currently on this site. We do suggest using the tools below to find Joshua Baldridge. You are visiting the placeholder page for Joshua Baldridge. This page is
Figure 6: Full version of Figure 4b: The joint distribution of the memorization score of each training example and its maximum influence on any validation set example. The histograms are in log scale to better visualize the tail of the distributions.

Example 2048352 (mem = 0.0320) link ▷ This is a placeholder page for Laytoya Brannon, which means this person is not currently on this site. We do suggest using the tools below to find Laytoya Brannon. You are visiting the placeholder page for Laytoya Brannon. This page is here because someone used our placeholder utility to look for Laytoya Brannon. We created this page automatically in hopes Laytoya Brannon would find it. If you are not Laytoya Brannon, but are an alumni of Mainland High School, register on this site for free now.

Example 1314053 (mem = 0.0278) link ▷ This is a placeholder page for Devin Mcguire, which means this person is not currently on this site. We do suggest using the tools below to find Devin Mcguire. You are visiting the placeholder page for Devin Mcguire. This page is here because someone used our placeholder utility to look for Devin Mcguire. We created this page automatically in hopes Devin Mcguire would find it. If you are not Devin Mcguire, but are an alumni of Kankakee Valley High School, register on this site for free now.

Example 1085524 (mem = 0.0209) link ▷ This is a placeholder page for Anthony Christie, which means this person is not currently on this site. We do suggest using the tools below to find Anthony Christie. You are visiting the placeholder page for Anthony Christie. This page is here because someone used our placeholder utility to look for Anthony Christie. We created this page automatically in hopes Anthony Christie would find it. If you are not Anthony Christie, but are an alumni of Old Bridge High School, register on this site for free now.

Figure 7: The joint distribution of memorization and simplicity. The histograms are plotted in log scale to better visualize the tail of the distributions.
Figure 8: Spearman’s R between memorization rankings from a set of \( m \) models and our full set of 400 models. As more models are trained, the ranking changes very little, with the ranking at 192 models having a Spearman’s R of at least 0.992 on all datasets.

Figure 9: The variance in memorization scores decreases significantly as the number of models increases for all 3 datasets.

Measurements of the edit distances show that they are near the boundary of the deduplication threshold chosen in Lee et al. [2021]. On the other hand, the tail of the distribution — examples with high counterfactual memorization are mostly unaffected by text deduplication.

D Variance of Memorization Scores

In Figure 8, we measure the Spearman’s R between our total set of 400 models and an \( m \) model subset. As expected, as \( m \) increases, so does Spearman’s R—in particular, at 192 models, the Spearman’s R is at least 99.2% for all datasets, and increasing \( m \) already appears to have diminishing returns.

Using the same partitioning into size \( m \) sets of models, we analyze the variance of memorization scores assigned to each sample. To do this, within each partition, we compute the memorization score assigned to each sample. We then compute the standard deviation of all partitions’ memorization scores for each sample. In Figure 9, we plot each sample’s standard deviation — in all, this demonstrates the distribution of the variance of memorization scores. We find that the variance decreases substantially as \( m \) grows, and concentrates near 0 already with \( m = 192 \), for all datasets.

E Histogram of Max-Influence on Generated Texts

Figure 10 shows the histogram of max-influence on each generated example by Grover-Mega (p=0.96) [Zellers et al., 2019], from the RealNews training examples. Those generated examples are publicly released at https://github.com/rowanz/grover/tree/master/generation_examples.

F Miscellaneous Experiment Details

Our experiments are implemented using JAX [Bradbury et al., 2018] and Flax [Heek et al., 2020], both open sourced library under the Apache-2.0 license. In the study of influence on generated texts, we use the publicly released generations from the Grover models [Zellers et al., 2019], available at their open source code repository, under the Apache-2.0 license.

We run the experiments using our internal cluster. The majority of the compute is consumed by model training. In this paper, we use standard training setup for transformer based neural language models,
Figure 10: Histogram of max-influence on each generated example by Grover-Mega (p=0.96), from the RealNews training examples.

Figure 11: Comparison of hash based subset sampling with numpy.random.choice.

which could run on single node machines with one or multiple GPUs. However, to carry out the full analysis, we need to train 400 different models for each of the three datasets analyzed in this paper.

G Subsampling Procedure

In the estimation of memorization and influence, we trained 400 models each on an independent random subset of training examples. We use Tensorflow Datasets (TFDS)\(^2\) to load our training data. TFDS supports loading a continuous range of examples, but does not support subset loading from a list of indices of individual examples. The API has a filter function which allows us to provide a Tensorflow predicate to precisely control the subset loading. However, a naive implementation of checking whether the index of the current example is in a given list of subset indices is very slow and scales poorly with the subset size.

To mitigate the issue, we implemented a hash based subset sampling predicate that can be evaluated efficiently for each example, and (approximately) select a random subset of a specified size. Let \(N\) be the total number of training examples, \(n < N\) be the expected subset size. The idea is to map the index \(i\) of each example to \(N/n\) hash buckets, and select all the examples that fall into one particular bucket. To make sure each model gets an independent subset sampling, we need to use different hash functions for different models. In our implementation, we compose a known hash function for uint64 types with a simple pseudo number based on the index of the current model to achieve this. Note the subset size sampled is close to \(n\) but is not guaranteed to be exactly \(n\). But this is not a problem in our settings. The specific implementation is shown below:

```python
def hash_sampler(mod, seed, system):
    """Get hash based subset sampler."
    Args:
        mod: total_n_egs // subset_size
        seed: different seed leads to different subset sample
```

\(^2\)https://www.tensorflow.org/datasets
system: 'np' or 'tf'.

Returns:
  A Tensorflow or Numpy subset sampler.
  
  np_hash = hash_uint64_builder('np')
  mul, offset, remainder = np_hash(seed + 1234 + np.arange(3))
  remainder = remainder % mod

  if system == 'np':
    def np_sampler(n_total):
      x = np.arange(n_total, dtype=np.uint64)
      return np_hash(x*mul + offset) % mod == remainder
    return np_sampler
  elif system == 'tf':
    tf_hash = hash_uint64_builder('tf')
    def tf_filter(idx, _):
      return tf.equal(tf_hash(idx*mul + offset) % mod, remainder)
    return tf_filter
  raise KeyError(f'Unknown system: {system}')

def hash_uint64_builder(system):
  """Build a hash function in tf/np for uint64."""
  if system == 'np':
    uint64_cast = functools.partial(np.array, dtype=np.uint64)
    op_xor = operator.xor
    op_rshift = operator.rshift
  elif system == 'tf':
    uint64_cast = functools.partial(tf.cast, dtype=tf.uint64)
    op_xor = tf.bitwise.bitwise_xor
    op_rshift = tf.bitwise.right_shift
  else:
    raise KeyError(f'Unknown system: {system}')

  # https://stackoverflow.com/questions/664014/
  # what-integer-hash-function-are-good-that-accepts-an-integer-hash-key
  def hash_uint64(x):
    x = uint64_cast(x)
    x = op_xor(x, op_rshift(x, 30)) * uint64_cast(0xbf58476d1ce4e5b9)
    x = op_xor(x, op_rshift(x, 27)) * uint64_cast(0x94d049bb133111eb)
    x = op_xor(x, op_rshift(x, 31))
    return x

  return hash_uint64

In Figure 11, we compare our hash-based subset sampler with numpy.random.choice(N, size=n, replace=False). The leftmost section of the figure shows that the sampling procedure always samples close to n points, with a small variance. The middle section plots a histogram of the empirical fraction of total models that each point appears in. Note that, because we use \( r = 0.25 \), this fraction should be 0.25 on average, although, because we only use 400 models, each value will not be identically 0.25. We find that our hash-based sampler produces probabilities which are highly consistent with those produced by numpy.random.choice. We also measure the pairwise independence of the hash-based sampler, measuring the probability that two different training points \( x_1, x_2 \) appear both IN or OUT of a model’s training set. We expect this value to be 0.625 (= \( r^2 + (1-r)^2 \)). We plot this in the right portion of the figure, demonstrating that the independence of our hash-based sampler is very similar to numpy.random.choice.
Alternative Memorization Metrics with Logit Scaling

We defined the counterfactual memorization in (1) with a generic performance measure $M$. Throughout the paper, we define $M$ as per-token accuracy—the fraction of the times the model assigns the highest score to the true next token in the sequence. The finite value range could cause unnecessary compression for values near the interval boundary. As a result, the resolution of memorization estimation is lower for models with very high or very low performance. To mitigate this issue, we explore an alternative measure by taking the logit on the per-token accuracy [Carlini et al., 2021]. The logit function maps to $(-\infty, \infty)$ before aggregating across independently trained models. Figure 12 compares the scatter plots of average performance on IN / OUT models measured by the logit scaled per-token accuracy and the raw per-token accuracy. Comparing to the raw per-token accuracy, the scatter plots generated with the logit scaled measure are no longer artificially constrained to be a triangular shape. As a result, the memorization estimation, which is proportional to the distance to the diagonal line, has a higher resolution on the two ends (lower left and upper right) than the unscaled version.

Note there is no absolutely right or wrong measure. While the scaled version has better resolution on the two ends, the advantage of the unscaled version is that the value range $[0, 1]$ makes it straightforward to interpret the numerical values of counterfactual memorization. Since the consistency between the two versions are high (Spearman’s $\rho$ correlation between the two versions are 0.947 / 0.903 / 0.944 on RealNews/ C4/ Wiki40B:en), we use the unscaled version throughout the paper for easier interpretation.

Definition of Edit Similarity

We define the edit similarity between two sequences $x_i$ and $x_j$ as. In our case, we use token-level similarity.

$$
\text{EditSim}(x_i, x_j) = 1 - \frac{\text{EditDistance}(x_i, x_j)}{\max(|x_i|, |x_j|)}
$$
Table 3: Pairs of RealNews training examples and Grover generations sampled at several influence levels. “link” contains the document URL. [ ] indicate text omitted for brevity. Differences in each pair are highlighted.

<table>
<thead>
<tr>
<th>Index</th>
<th>Estim.</th>
<th>Text</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generation 1361</td>
<td>infl</td>
<td>[link] Baku, Azerbaijan. April 15 Trend: Official exchange rate of the US dollar and euro against Azerbaijani manat was set at 1.7 and 1.92 manats, respectively, for April 15. Below are the rates of Azerbaijani manat against world currencies, according to the data from the Central Bank of Azerbaijan for April 15. [ ] 100 Japanese yen 100 JPY 1.51 Follow Trend on Telegram. Only most interesting and important news</td>
</tr>
<tr>
<td>Train 2072973</td>
<td>mem</td>
<td>[link] Baku, Azerbaijan, March 15 Trend: Official exchange rate of the US dollar and euro against Azerbaijani manat was set at 1.7 and 1.92 manats, respectively, for March 15. Below are the rates of Azerbaijani manat against world currencies, according to the data from the Central Bank of Azerbaijan for March 15. [ ] 100 Japanese yen 100 JPY 1.52 Follow Trend on Telegram. Only most interesting and important news</td>
</tr>
<tr>
<td>Generation 21998</td>
<td>infl</td>
<td>[link] NEW DELHI: India is likely to receive average monsoon rains this year, the state-run weather office said on Monday, which should support agricultural production and economic growth in Asia’s third-biggest economy, where half of the farmland lacks irrigation. Monsoon rainfall is expected to be 99 percent of the long-term average, M. Rajeevan, secretary at the Ministry of Earth Sciences, told a news conference. The India Meteorological Department (IMD) defines average, or normal, rainfall as between 96 percent and 104 percent of a 50-year average of 89 centimeters for the entire four-month season beginning June. [ ] India’s weather office will update its forecast in the first week of June. However, on average, the IMD has forecast accurately only once every five years over the past two decades, even after taking into account an error band of plus or minus 5 percentage points.</td>
</tr>
<tr>
<td>Train 326212</td>
<td>mem</td>
<td>[link] NEW DELHI (Reuters) - India is likely to receive average monsoon rains in 2018, the weather office said, raising the possibility of higher farm and economic growth in Asia's third-biggest economy, where half of the farmland lacks irrigation. Monsoon rainfall, the lifeblood of the country’s $2.6 trillion economy, are expected to be 99 percent of a long-term average, K.J. Ramesh, director general of the state-run India Meteorological Department (IMD), told a news conference. “We see very less probability of a deficit monsoon,” Ramesh said on Monday. Other than lifting farm and wider economic growth, a spell of good rains will keep a lid on inflation, potentially tempting Prime Minister Narendra Modi to bring forward general elections due in May 2019. India’s weather office defines average, or normal, rainfall as between 96 percent and 104 percent of a 50-year average of 89 cm for the entire four-month season beginning June. [ ] said a Mumbai-based dealer with a global trading firm. Average monsoon rainfall will help India retain its position as the world’s top rice exporter.</td>
</tr>
</tbody>
</table>

J Examples of Train-Generation Pairs at Different Influence Ranges

In table 3, we show examples of train-generation pairs sampled from different influence ranges. The patterns generally follow the train-validation pairs shown in table 2, although many of the relations are due to some form of templating.

K Examples Sampled at Different Level of Memorization

Figure 13, Figure 14, and Figure 15 show full examples from RealNews sampled at high, middle and low memorization value ranges, respectively. Similarly, Figure 16, Figure 17, and Figure 18 show examples from C4 sampled at high, middle and low memorization value ranges, respectively. Figure 19, Figure 20, and Figure 21 show examples from Wiki40B:en sampled at high, middle and low memorization value ranges, respectively.

L Example Pairs Sampled at Different Level of Influence

Figure 22, Figure 23, Figure 24, Figure 25, and Figure 26 show train-validation example pairs from RealNews sampled from high to low influence ranges. For each pair, we show the validation set example first, and then show the corresponding training example with a diffLib generated visualization of textual difference with the training example.

Similarly, Figure 27 and Figure 28 show train-validation example pairs from C4, and Figure 29 and Figure 30 from Wiki40B:en.

We also show train-generation influence pairs between RealNews training set and Grover [Zellers et al., 2019] model generation in Figure 31, Figure 32, and Figure 33.
I would like to acknowledge that this meeting is being held on the traditional lands of the (appropriate group) people, and "[local Indigenous group's name] I welcome you all. [insert name of Indigenous people], the traditional custodians of this land where we are meeting upon today. On behalf of the traditional custodians representative/Elder of the [insert organisation or local Indigenous group]. I would like to begin by paying my respect to the local Indigenous people [or the] [indigenous group] and their elders past, present and future. In the spirit of [insert local Indigenous language] I give thanks for the land."
A Texas honors student punished for saying that homosexuality was wrong has had his suspension rescinded after a meeting with the mother and her attorney. The 14-year-old, Dakota Ary, from Western Hills High School of the Fort Worth Independent School District was initially given a suspension of one day in-school and two days full. After hearing about the suspension the mother, Holly Pepe, reached out to Matt Krause of the Liberty Counsel to be her son’s legal representative. Matt Staver, founder and chairman of the Liberty Counsel, told The Christian Post that he believed Western Hills High made the correct decision in reversing their course of action. “The decision to rescind the suspension is the correct one. The suspension was wrong and improper,” said Staver. “I applaud the student for standing up. We stood with him to resist an unjust suspension and we are pleased that suspension has been reversed.” Dakota Ary was in a German language class at Western Hills when class conversation shifted to the issues of religion and homosexuality. The student reportedly said that “being a homosexual was wrong.” Upon hearing his remark, the teacher proceeded to take administrative action. Ary himself is a non-student, noted for being a member of the high school’s football team and a volunteer at the school’s track team. The suspension allows for him to play in an upcoming football game. In a publicly released statement, the Fort Worth Independent School District said that it does not comment on student matters. In keeping with this, Fort Worth ISD did not return comment to The Christian Post by press time. “As a matter of course, Fort Worth ISD does not comment on specific employee or student-related issues. Suffice it to say that we are following district policy in our review of the circumstances and any resolution will likewise be in accordance with district policy,” said Fort Worth ISD. Staver also told CP that in recent years incidents like this one are only increasing as the gay rights movement continues to advance in American society. “Unfortunately, we have seen similar situations. The homosexual agenda is aggressive and intolerant to people who do not agree. Liberty Counsel will continue the right to exercise freedom of conscience and religion,” said Staver. “These instances are increasing and will continue to increase unless Christians and people who love liberty stand up and resist this intolerance.”

Stella & Dot, the social commerce-based accessories company, will add to its offerings on April 8 with a full range of handbags. The nine-piece collection will retail from $21 for a cosmetics case to $138 for a weekend bag. Founder and chief executive officer Jessica Herrin called this the nearly 10-year-old company’s “launch into a lifestyle brand,” as well as a chance to double its market opportunity. “In the $10 billion accessories market in the U.S., bags and jewelry each account for approximately $10 billion,” Herrin told WWD. “We see the same opportunity in bags as we did in jewelry — giving women on the go a simple way to be chic.” Other styles include a $59 technology case that serves double duty as a clutch and wallet, a $99 jewelry roll, a classic $89 tote and a $128 convertible bag that has zippers up the side to let the wearer decide if she wants a sleeker or more expanded silhouette. A small handbag collection that launched in 2011 — containing two cross-body and two convertible cross-body to clutch styles — but this is the first significant push for handbags geared for daytime use. With sales expected to surpass $200 million this year — business grew from $175 million to $200 million from 2011 to 2012 — Herrin projected the new category will take in about $25 million through the end of the year. In addition to the brand being carried exclusively online at stelladot.com, the brand is sold via 14,000 active stylists around the world. “We’re really invested in this as a huge growth category. We just opened a dedicated in-house design studio for this in Sausalito (Calif.),” Stella & Dot chief creative officer Blythe Harris said. She added that the bags are comprised of a combination of coated canvas, waffle poly and Saffiano leather and come in black, snakeskin and colorful multistripe prints. Herrin added: “Broadening our accessories line by launching the new and massive category of bags allows us to make that opportunity for stylists bigger than ever before. It’s a ground-floor opportunity with a proven company.”

Figure 14: Text examples from RealNews with intermediate memorization.
Figure 15: Text examples from RealNews with low memorization.
Color Enhancing Blood Orange & Passion Fruit Shampoo enhances red tones on red, light brown, and dark blonde hair. Apply shampoo to wet hair. Lather, leave on 3-5 minutes & rinse. Water/Aqua/Eau, Sodium Laureth Sulfate, Cocamide MIPA, Cocamidopropyl Betaine, Glycol Stearate, Citrus Aurantium Dulcis (Orange) Fruit Extract, Passiflora Edulis Fruit Extract, Rubus Nigricans (Black Currant) Fruit Extract, Lycium Barbarum Fruit Extract, Cocos Nucifera (Coconut) Oil, Argania Spinosa Kernel Oil, Gardenia Tahitensis Flower Extract, Laureth-12, Laureth-23, Guar Hydroxypropyltrimonium Chloride, PEG/PPG-4/12 Dimethicone, Polyquaternium-11, Citric Acid, Tetrasodium EDTA, PEG-150 Pentaeerythritol Tetrastearate, PEG-6 Caprylic/Capric Glycerides, Ethylhexyl Salicylate, Ethylhexyl Methoxycinnamate, Butyl Methoxydibenzoylmethane, PEG-6 Hydrogenated Castor Oil, Glycerin, PPG-26-Buteth-26, Polysorbate 20, Imidazolidinyl Urrea, Limonene, Fragrance/Parfum, Orange 4 (CI 15510), Red 33 (CI 17200).

Figure 16: Text examples from C4 with high memorization.
Citizen scientists from across Australia and the world have helped researchers count Illawarra’s elusive spotted-tailed quoll, Environment Minister Gabrielle Upton said 3 March. “Six months ago we asked citizen scientists to jump online and help us analyse more than 80,000 photos taken from across the Illawarra region,” Ms Upton said. “More than 300 volunteers answered the call and we can now confirm at least 20 individual quolls call this region home. “This is a great milestone for this citizen science project – in time to celebrate World Wildlife Day!” The Quollidor project is funded under the NSW Government’s $100 million Saving our Species project and is providing a fascinating glimpse into the hidden world of this endangered carnivorous marsupial. The 29 motion sensor cameras in the region collect 20,000 to 30,000 images every 8 weeks that the enthusiastic citizen scientists have sorted through. “We have images of quolls posing for selfies, exploring the camera and jumping on and around the monitoring station,” Ms Upton said. Quolls have huge home ranges and move across the landscape making it difficult to otherwise monitor their population and behavioural patterns. “It’s great to see people being able to use technology to make a real contribution to the conservation of NSW’s unique animals,” Ms Upton said. The spotted-tailed quoll is the only remaining quoll species in the state. The project is ongoing and will help the NSW Government increase the resilience and size of the local quoll population.

Yesterday’s trip to our local, Le Chatelain, was a welcome reprieve, and I took the opportunity to sample a new beer, Floreffe Blond. Hard to believe I haven’t had everything on offer there yet, but that’s Belgian pulls for you: even a run-of-the-mill one will have 30 or so types. Le Chatelain probably has closer to 60. Floreffe’s dark beers, Floreffe Dubbele and Floreffe Prima Meller, are included in my list of top ten Belgian beers, so I was expecting great things with Floreffe Blond. I wasn’t disappointed. Floreffe Blond is lighter than normal strong blond ales. It’s lemony, summery. Fiona says it’s smooth and has a bubble gum taste -- “that powdery, old-style bubble gum” -- but I’m sticking with citrus. It has very little bitterness, but it’s not sweet. At 6.3% alcohol, it’s not too brain-bending either. Which is just as well, because the wood treatment is doing my head in plenty enough. 

The International Republican Institute (IRI) has deployed a Long-Term Observation Mission to observe preparations for the Moldovan parliamentary elections scheduled for February 26, 2019. Long-term observers (LTOs), who were accredited by the Moldovan Central Election Commission on December 11, 2018, are based in Edinet, Ungheni, Orhei, Anenii Noi, Hancesti, Comrat and Chisinau. IRI observers will remain on the ground to observe the pre-election preparations, Election Day and the post-election period, culminating their mission on March 26, 2019. The mission is led by Andrea Keerbs, with the assistance of three analysts who will provide in-depth analysis of the media, electoral and legal landscape. The LTOs represent 11 different countries: the U.S., Poland, Canada, the U.K., France, the Philippines, Uganda, Portugal, Montenegro, Slovenia and Georgia. IRI has organized more than 200 international election observation missions around the globe, earning a reputation for impartiality and professionalism. The IRI mission to Moldova is funded by the United States Agency for International Development and will conduct its activities on a strictly independent and nonpartisan basis, without interfering in the election process and conformity to the laws of Moldova. Click here for more information on IRI’s work in Moldova.

Figure 17: Text examples from C4 with intermediate memorization.
This is a placeholder page for Lewis Thompson, which means this person is not currently on this site. We do suggest using the tools below to find Lewis Thompson. You are visiting the placeholder page for Lewis Thompson. This page is here because someone used our placeholder utility to look for Lewis Thompson. We created this page automatically in hopes Lewis Thompson would find it. If you are not Lewis Thompson, but are an alumni of Odessa High School Odessa, MO, register on this site for free now.

This is a placeholder page for David Baldwin, which means this person is not currently on this site. We do suggest using the tools below to find David Baldwin. You are visiting the placeholder page for David Baldwin. This page is here because someone used our placeholder utility to look for David Baldwin. We created this page automatically in hopes David Baldwin would find it. If you are not David Baldwin, but are an alumni of Anaheim High School, register on this site for free now.

At Shower Replacement Guys, our experts understand the process of Shower Replacement. This is an important training that ensures that your tub or shower so that it discards the old, worn out and outdated look for a breathtaking and refreshing replacement shower doors. The process starts with removing panels, loosening the jambs, removing the bottom track, installing the side tracks and leveling the track. They will then expertly fasten the jambs, expand the holes for the panels, installing back the bottom track and the storage column. As the process nears the end stage, our experts will then fasten the jambs to the column before restoring the top track. Once the shelving units are expertly restored, the caulk frames will be put in place and finally, you will have the door back in place. Talk to our experts in Shower Replacement Guys in South Gibson, PA so that you can learn more on how all these stages can be carried out efficiently and conveniently for your Shower Replacement needs. One of the key requirements for quality services in our South Gibson, PA is that the services must be sanctioned by the authorities. We uphold provision of quality services that meet the needs and requirements of people so that they can enjoy services provided without fear of substandard services. We have complied with all the provisions of the law so that you get superior quality services. As a result, you can be sure that compare favorably with other providers of Shower Replacement services in South Gibson, PA. You can therefore approach us for our services with full confidence that you will get the quality services. We guarantee you that you will get the most trusted and reliable shower door replacement services from Shower Replacement Guys in South Gibson, PA. This replacement shower doors will transform your experience with your shower or tub. It will spot a new look replete with elegance, style and an air of sophistication never experienced before. Call us on 888-398-0573 and discover the extent to which your old tub r shower will be changed from the unpleasant look to a most appealing and attractive finish unprecedented.

At Shower Pan Guys, you will find that our personnel have the requisite training, skills and expertise in the design of the Shower Pan liner or Shower Pan. Whether you are looking for a flexible shower liner, PVC,One liner, prefabricated shower liner, roll on membrane among other options, you will be at home with our experts to advice you on which Shower Pan is most suitable for your shower. There are some specific details that only professionals can handle. The reason why this issue is important is that you would to use shower liner for water proof environment or other flexible options available. You will require people with expertise to guide you on all these. Thank fully, we do not run out of these exceptionally skilled people to assist you in making the right decision. At Shower Pan Guys in Greenbush, WI, all your shower liner needs will be catered for in a most professional way. Our personnel will implement your proposals in an effortless way. You can be guaranteed of fully bonded, insured and licensed service in Greenbush, WI. We take our work seriously and therefore ensure that all the requirements as set by the local government are adhered to. The safety of any installation is as good as the installer. We do not like putting our clients at a disadvantage with leaking Shower Pan liner such that they cannot enjoy their bathroom experience. As per set rules and regulations, we follow the right procedure in ensuring that you get the full worth of your investment from procurement to the final installment stage. We also rank favorably with other similar service providers in Greenbush, WI. Have the confidence to approach at any time for a quality and high standard services. You will find that from Shower Pan Guys in Greenbush, WI, we help alleviate the shipping cost for our customers. For purchases worth some predetermined amount, you will receive shipping to your destination. Call us on 888-670-3340 and learn more the amount of purchases that will enable you to receive free shipping services. Shower Pan Guys in Greenbush, WI can be contacted over 888-670-3340 to have any aspect of the shower pans clarified to you in way you will understand. Issues such as procurement, quotation, design, inspection, installation, repair and maintenance will be explained to you so that you understand the procedure. You will be able to reach us through email or the online option. Our personnel will be very helpful to you so that you can make the right decision regarding your requirements.
Figure 19: Text examples from Wiki40B:en with high memorization.
In March 2010, the 6-year trial ended with a verdict of not guilty. The judge in the case noted the pilot expertly carried out the difficult autorotation landing and that the passengers survived because of his superb piloting skills. Leszek Miller declared that if he had to fly again in a helicopter in difficult atmospheric conditions, he would choose Milosz as his pilot.
Eleven men went down with the U-boat; there were 45 survivors. The boat departed Kiel on 15 April 1943, moved through the North Sea, negotiated the gap between Iceland and the Faroe Islands, and a 3.7 cm (1.5 in) C/30 anti-aircraft gun. The boat had a complement of between forty-four and sixty.

**START SECTION** Service history

The boat's service life began with training with the 8th U-boat Flotilla in October 1943. She was reassigned to the 13th U-boat Flotilla for operations on 1 August 1944. She was reassigned to the 9th Flotilla for operations on 1 August 1944. She was reassigned to the 5th Flotilla on 1 August 1943. The boat's service life began with training with the 8th U-boat Flotilla in October 1943. She was reassigned to the 13th U-boat Flotilla for operations on 1 August 1944. She was reassigned to the 9th Flotilla on 1 October and moved again to the 14th Flotilla on 1 April 1945.

**START SECTION** Design

**START ARTICLE** German submarine U-295 _START PARAGRAPH_ German Type VIII/41 submarines were preceded by the shorter Type VIII submarines. U-295 had a displacement of 759 tonnes (747 long tons) when at the surface and 860 tonnes (850 long tons) while submerged. She had a total length of 67.30 m (220 ft 2 in), a pressure hull length of 50.50 m (166 ft 8 in), a beam of 6.20 m (20 ft 4 in), and a draught of 4.74 m (15 ft 6 in). The submarine was powered by two Germaniawerft F46 four-stroke, six-cylinder supercharged diesel engines producing a total of 2,800 to 3,200 metric horsepower (2,060 to 2,350 kW; 2,760 to 3,160 shp) for use while surfaced, two AEG GU 460/8-27 double-acting electric motors producing a total of 750 metric horsepower (550 kW; 740 shp) for use while submerged. She had two shafts and two 1.23 m (4 ft) propellers. The boat was capable of operating at depths of up to 230 metres (750 ft). **NEWLINE** The submarine had a maximum surface speed of 17.7 knots (20.4 mph) and a maximum submerged speed of 7.6 knots (8.7 mph). When submerged, the boat could operate for 80 nautical miles (150 km; 92 mi) at 4 knots (7.4 km/h; 4.6 mph); when surfaced, she could travel 8,500 nautical miles (15,940 km; 9,800 mi) at 10 knots (19 km/h; 12 mph). U-295 was fitted with five 53.3 cm (21 in) torpedo tubes (four fitted at the bow and one at the stern), fourteen torpedoes, one 8.8 cm (3.46 in) SK C/35 naval gun, (220 rounds), one 3.7 cm (1.5 in) Flak M2 and two 2 cm (0.79 in) C/70 anti-aircraft guns. The boat had a complement of between forty-four and sixty.

**START SECTION** 5th patrol

Her fifth effort was just as barren, even though it was longer. **NEWLINE** Her fifth patrol started in Harstad and finished in Narvik. She had spent three days off Murmansk, to no avail. **NEWLINE** Then she embarked on a series of short journeys between Bergen, Kristiansand, Stavanger, and Trondheim, NEWLINE, her second foray, between Trondheim and Harstad was the most successful. She damaged the British frigate HMS Mouseyeast northeast of Murmansk on 2 November 1944, **NEWLINE** Her fourth sortie took her into the Barents and Norwegian Seas. She returned to Harstad on 18 December 1944, **NEWLINE** Her fourth patrol started in Harstad and finished in Narvik. She had spent three days off Murmansk, to no avail. **NEWLINE** Her fourth sortie took her into the Barents and Norwegian Seas. She returned to Harstad on 18 December 1944. **NEWLINE** Her fifth patrol **NEWLINE** Her fifth patrol started in Harstad and finished in Narvik. She had spent three days off Murmansk, to no avail.

**NEWLINE** Her fifth patrol _START PARAGRAPH_ The boat departed Narvik on 15 April 1945. Her route took her once again to the Barents Sea. She returned to the Nordic port on 7 May, **NEWLINE** She was then moved to Skijómenford on 12 May 1945 and in accordance with the surrender terms, she was transferred to Loch Erriboll in northern Scotland for Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

**NEWLINE** Operation Deadlight on the 19th. She was sunk on 17 December by the guns of ORP Blyawska.

Figure 21: Text examples from Wiki40B:en with low memorization.
By Ari Rabinovich TZEELIM VALLEY, Israel (Reuters) - The disposable paper face masks offer little protection from the clouds of dust that fill the cliffside cave where Israeli archaeologists are wrapping up the largest excavation in the Judean desert of the past half-century. Clipped into safety harnesses, volunteers stand at the cave opening, 250 meters (820 feet) above a dry river bed that leads to the lowest spot on earth, the Dead Sea. They sift through an endless supply of dirt-filled buckets, and the dust they throw in the air reaches the far corners of the cave where a dozen workers crawling on hands and knees can't help but cough. The three-week excavation was the first part of a national campaign to recover as many artefacts as possible, particularly scrolls, left behind by Jewish rebels who hid in the desert some 2,000 years ago, before they were snatched up by antiquity robbers. "These looters that operate in the area are experts at finding scrolls. We go after them, look for what they are looking for and try to catch them," said Guy Fitoussi, head of the Israel Antiquities Authority robbery prevention unit in southern Israel. "This is the game. Like cat and mouse." The Dead Sea Scrolls, a collection of ancient texts written on papyrus and parchment, have already been rescued by scholars. They are among the earliest texts written in the Hebrew language and are on display in the Israel Museum in Jerusalem as a national treasure. Now Israel wants to uncover whatever may remain in the desert hideouts before it is destroyed or ends up on the black market. For a Wider Image photo essay on the excavation click: http://reut.rs/25zc4ZK

According to Israeli law, all relics found on land or at sea belong to the state. Fitoussi, a pistol-packing archaeologist with authority to arrest looters, and his team catch about 100 of them each year. Most are fined; some are sent to jail. In 2014, they arrested six people who were plundering this particular cavern, known as the Cave of Skulls, where seven skulls had been found from Jews of the Bar Kokhba rebellion against Rome in the 2nd century. That raid, Fitoussi said, helped spur the multi-year government-backed excavation program and focused their initial efforts at this site, about a two-hour drive southeast from Jerusalem. To access the cave, diggers don climbing gear and descend 20 minutes from their campsite along a steep path that hugs the rocky cliff. Inside, the grotto expands 160 square meters (1,720 square feet), including a number of cramped tunnels that extend deep into the mountain. The limestone walls of the dry desert cave are perfect for preservation, said Uri Davidovich, an archaeologist from Tel Aviv University who was one of the dig's directors. One 19-year-old volunteer, working flat on her belly in a dark crawl space, dirt mixed with sweat covering her face and digging with her fingers, unearthed a thin, 25 centimeter (10 inch)-long rope that most likely was used by the Bar Kokhba rebels. A rope this length was a rare discovery, Davidovich said. They haven't found any scrolls yet, he said, but the artefacts found in this cave, and countless others nearby, will provide historians rare insight into how people lived 2,000 to 8,000 years ago. (Editing by Jeffrey Heller/Jeremy Gaunt)
VATICAN CITY (AP) — Emeritus Pope Benedict XVI is offering a first-ever papal assessment of his own pontificate in a book that recounts his decision to resign, his surprise at his successor and his attempts to dismantle what he calls the Vatican’s “gay lobby.” "Benedict XVI: The Final Conversations," is due out in September, the latest book-length interview that Benedict has conducted with German journalist Peter Seewald. Italian daily Corriere della Sera, which has the book's newspaper rights, provided a brief overview Friday. Corriere said Benedict recounts in the book that he decided to announce his resignation in Latin because he feared making a mistake in Italian. He recalls his “surprise” that Jorge Mario Bergoglio was elected pope and his “joy” at seeing Pope Francis mingle with crowds. Benedict also claims to have dismantled a group of four or five gay prelates, dubbed the “gay lobby” by the Italian media, who exercised power and influence in the Vatican. The existence of this group of gay prelates — who purportedly used blackmail to promote and preserve their interests — has been mythologized in Italian media. Soon after he was elected pope and was asked about the so-called “gay lobby,” Francis quipped that he had yet to encounter any priest who had “gay” written on his business card. That said, just this week a gay monsignor who was fired from the Vatican and suspended as a priest after he came out, boyfriend by his side, published a book about his experiences as a gay official in the Vatican's doctrine office. In "The First Stone," Polish-born Krzysztof Charamsa recounts the absolute “obsession” with homosexuality in the halls of the Holy See. He details the “hypocrisy” of its functionaries who profess a celibate life but live quite another, and writes that it was enough to destroy someone's Vatican career by simply spreading gossip that he was gay.

Figure 23: Validation / training example pair from RealNews with relatively high influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
ANAHEIM – On a night when Francois Beauchemin had two assists in Toronto, and Chris Pronger blocked six shots in Detroit, 13,869 Ducks fans might have been lost without their programs on Wednesday night. It's only the first game of the pre-season, but a new era has clearly begun. A group of mostly newcomers in Ducks uniforms beat Phoenix 3-2 in a shootout on Wednesday at Honda Center. A familiar face made the biggest impact, however, as Bobby Ryan scored two goals. Ryan also scored two goals in the Ducks’ pre-season opener last year, when he was trying to make the team’s opening-day roster as a rookie. His bid failed and Ryan was forced to start the season in the American Hockey League, mostly due to salary-cap constraints. The circumstances could not be much different this year. Ryan finished the season as a Calder Trophy candidate, and began this season wearing a temporary alternate captain’s “A” on his jersey along with Joffrey Lupul and Sheldon Brookbank. Nick Boynton, a 30-year-old defenceman signed in the off-season, was the eldest player in uniform. Of the 20 players chosen to play, only goaltender Jonas Hiller and center Ryan Carter spent all of last season with the Ducks. Hiller started in goal and looked sharp, stopping 18 of the 19 shots he faced. Timo Pielmeier, acquired in the trade that sent Kent Huskins and Travis Moen to San Jose, came in halfway through the third period and stopped 27 of 29 shots before the game went into overtime. A phantom interference penalty sent Ducks defenceman Steve Eminger to the box, and set up a power-play goal by the Coyotes’ Lauri Korpikowski with 20.2 left in the period. Korpikowski deflected a long-range shot by Keith Yandle, and Hiller did well just to get his glove on it. Ryan got the goal back early in the second period, poking in a rebound of a Dan Sexton shot past former Kings goalie Jason LaBarbera. With 1:08 elapsed in the third period, Ryan made it 2-1 on a breakaway shot that caromed in off a Phoenix stick. Phoenix tied the game at 2 at 12:30 of the third period, when Sheldon Brookbank made a costly turnover in front of his own net. Chad Kolarik batted down Brookbank’s pass just a few feet in front of Pielmeier and tucked the puck in behind the goaltender.

Figure 24: Validation / training example pair from RealNews with intermediate influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
More than 70,000 pounds of Butterball turkey recalled because of potential salmonella

WASHINGTON — The U.S. Department of Agriculture’s Food Safety and Inspections services announced on Wednesday that approximately 78,164 pounds of raw ground Butterball turkey might be contaminated with Salmonella. Product numbers "EST. P-7345" inside the USDA mark of inspection are subject to recall. They were shipped to nationwide retail and institutional locations.

The illness lasts four to seven days. According to the USDA, most people recover without treatment, however, they suggest that if diarrhea is severe the person might need to be contaminated. They also suggest those with weakened immune systems are more susceptible to illness. Consumers with food safety questions can "Ask Karen," the FSIS virtual representative available 24 hours a day at AskKaren.gov or via smartphone at m.askkaren.gov.

The toll-free USDA Meat and Poultry Hotline 1-888-674-6854 is available in English and Spanish and can be reached from 8 a.m. to 4 p.m. MDT Monday through Friday. Recorded food safety messages are available 24 hours a day. The online Electronic Consumer Complaint Monitoring System can be accessed 24 hours a day. RELATED: USDA RECALLS

More than 70,000 pounds of Butterball turkey recalled because of potential salmonella

WASHINGTON — The U.S. Department of Agriculture’s Food Safety and Inspections services announced on Wednesday that approximately 78,164 pounds of raw ground Butterball turkey might be contaminated with Salmonella. Product numbers "EST. P-7345" inside the USDA mark of inspection are subject to recall. They were shipped to nationwide retail and institutional locations.

The illness lasts four to seven days. According to the USDA, most people recover without treatment, however, they suggest that if diarrhea is severe the person might need to be contaminated. They also suggest those with weakened immune systems are more susceptible to illness. Consumers with food safety questions can "Ask Karen," the FSIS virtual representative available 24 hours a day at AskKaren.gov or via smartphone at m.askkaren.gov.

The toll-free USDA Meat and Poultry Hotline 1-888-674-6854 is available in English and Spanish and can be reached from 8 a.m. to 4 p.m. MDT Monday through Friday. Recorded food safety messages are available 24 hours a day. The online Electronic Consumer Complaint Monitoring System can be accessed 24 hours a day. RELATED: USDA RECALLS

Figure 25: Validation / training example pair from RealNews with relatively low influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
Earlier in January, Satyapal Singh had claimed that Charles Darwin's theory of evolution of man was 'scientifically wrong' and it needed to be changed in school and college curriculum. (Photo: File) New Delhi: Minister of State for Human Resource Development, Satyapal Singh speaking at a meeting of the Central Advisory Board of Education (CABE) on January 15 and 16 said that mantras codified the 'laws of motion' much before it was discovered by Isaac Newton. He also suggested that Vaastu compliance of educational buildings was important for learning, according to a report in Hindustan Times. "There are mantras which codified 'laws of motion' much before it was discovered by the Newton. Hence it is essential that traditional knowledge must be incorporated in our curriculum," Singh was quoted as saying by the minutes of the meeting of the government's highest advisory body for policymaking in education.Earlier in January, Satyapal Singh had claimed that Charles Darwin's theory of evolution of man was "scientifically wrong" and it needed to be changed in school and college curriculum. Singh said our ancestors have nowhere mentioned that they saw an ape turning into a man.

President Hosni Mubarak said he is willing to visit Israel if the Jewish state shows flexibility in promoting peace and agrees to take part in an international Middle East peace conference, according to an interview published Saturday. "I am still saying that I am ready to travel to Israel provided that this would lead to real progress on the road to solving the Middle East problem," Mubarak told the Egyptian daily Al Ahram. The offer represented the second time in less than a week that Mubarak had expressed readiness to make a peace visit to Israel. In an interview with the Kuwaiti newspaper Anba, Mubarak said he was prepared to travel to Israel if the visit would promote a solution of the Palestinian problem and a just Middle East peace. The offer was initially embraced by Israel, but later two of Mubarak's senior foreign policy aides said the notion of a visit was contingent on the Jewish state opening negotiations with the Palestine Liberation Organization. Israel rejects the PLO as a negotiating partner despite the United States' scientifically wrong' and it needed to be changed in school and college curriculum. Singh said our ancestors have nowhere mentioned that they saw an ape turning into a man.

Figure 26: Validation / training example pair from RealNews with low influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
<table>
<thead>
<tr>
<th>Index</th>
<th>Estim.</th>
<th>Link and Text</th>
</tr>
</thead>
</table>

Range Repair 92811 Expert repairs and services all types ranges, whether you are in need of electric range repair or gas range repair. If your range is having problems like the range surface Element won't work, range burner has spark problems, range surface element won't turn off, range burners spark all the time. Range Repair Service will put you right back where you need to be. Our range repair technicians carry most range parts. Expert Appliance Repair 92811 will have your range repaired or serviced in no time flat. We will have your range up in running in no time and you back to cooking.


Range Repair 92817 Expert repairs and services all types ranges, whether you are in need of electric range repair or gas range repair. If your range is having problems like the range surface Element won't work, range burner has spark problems, range surface element won't turn off, range burners spark all the time. Range Repair Service will put you right back where you need to be. Our range repair technicians carry most range parts. Expert Appliance Repair 92817 will have your range repaired or serviced in no time flat. We will have your range up in running in no time and you back to cooking.

| valid | 274413 | http://vannafrosoni.featuredwebsite.com/free_home_valuation.asp |

Want to know what your home in ENGLISHTOWN is worth? I'll provide you with a FREE home evaluation at no obligation. Simply fill in your information below and I'll get back to you shortly with a detailed report of comparable homes in your area that have recently sold or are currently for sale.

| train | mem | 28486 | http://cnyrealestate.com/free_home_valuation.asp |

Want to know what your home in ENGLISHTOWN is worth? I'll provide you with a FREE home evaluation at no obligation. Simply fill in your information below and I'll get back to you shortly with a detailed report of comparable homes in your area that have recently sold or are currently for sale.

| valid | 27342 | https://www.sciway.net/hotels/beaches.html |

South Carolina SC Hotels SC Beach Hotels Also see: SC Coastal Hotels | US Hotels Want to stay at the beach? Below is an accommodation list for all the beaches in South Carolina – plus descriptions of what those beaches are like. Find hotels located on SC beaches, including Myrtle Beach, shown here. Our list starts at South Carolina's border with North Carolina and heads south towards Savannah, Georgia.

| train | mem | 1520782 | https://www.sciway.net/hotels/coast.html |

South Carolina SC Hotels SC Coastal Hotels Also see: SC Beach Hotels | US Hotels Want to stay at the beach along the coast? Below is an accommodation list for all the beach coastal areas in South Carolina – plus descriptions of what those beach areas are like. Find hotels located on SC beaches along the SC coast, including Myrtle Beach, Charleston, shown here. Our list starts at South Carolina's border with North Carolina and heads south towards Savannah, Georgia.

Figure 27: Validation / training example pair from C4 with high to intermediate influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
A Community Trigger gives victims and communities the right to demand action on problems with anti social behaviour (ASB) they have reported in the past. The Community Trigger can be used by anyone who has reported ASB but feels no action has been taken. The Community Trigger is aimed at putting victims first and to hold agencies responsible for managing anti social behaviour to account. Agencies including councils, the police, local health teams and registered providers of social housing who receive a Community Trigger report will then need to conduct a case review. Who can raise a Community Trigger and when? Anyone can raise the trigger on behalf of the victim - for example a family member, friend, care, councillor, Member of Parliament or other professional person. It doesn’t matter who you originally reported the ASB to (the council, the police or your landlord) - please use the Community Trigger form here on our website.

What happens after a Community Trigger is raised? When you complete the Community Trigger form, we will contact you to say we have received it and let you know what will happen next. Your completed form will be sent to Sandwell Council’s Community Safety and Anti Social Behaviour Manager. A member of the ASB team will contact you, within two working days. The Community Trigger is not an alternative method of complaining about the service you have received. If you are not satisfied with the service you have received there is still an independent complaints process that you should follow. Feedback on the use of the Community Trigger.

This is a 1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring. Diamond - Old Mine Cut - GH VS1-VS2. Item has been tested and guaranteed to be solid PLATINUM. This ring is in excellent pre-owned condition. Please disregard the two characters at the end of the title. You are using for inventory purposes. As a courtesy, please notify us of any return. Always fast & free unless otherwise stated. Collectors Coins & Jewelry has been family owned and operated on Long Island, NY since 1946. We have four brick and mortar locations and offer the highest quality products with unbeatable customer service. The item “1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring” is in sale since Thursday, November 30, 2017. This item is in the category “Jewelry & Watches\Jewelry\Gold & Watches & Brooches\Diamond & Gemstones”. The seller is “collectorsbuysell” and is located in Huntington, New York. This item can be shipped worldwide.

This is a 1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring. Diamond - Old Mine Cut - GH VS1-VS2. Item has been tested and guaranteed to be solid PLATINUM. This ring is in excellent pre-owned condition. Please disregard the two characters at the end of the title. You are using for inventory purposes. As a courtesy, please notify us of any return. Always fast & free unless otherwise stated. Collectors Coins & Jewelry has been family owned and operated on Long Island, NY since 1946. We have four brick and mortar locations and offer the highest quality products with unbeatable customer service. The item “1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring” is in sale since Thursday, November 30, 2017. This item is in the category “Jewelry & Watches\Jewelry\Gold & Watches & Brooches\Diamond & Gemstones”. The seller is “collectorsbuysell” and is located in Huntington, New York. This item can be shipped worldwide.

This is a 1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring. Diamond - Old Mine Cut - GH VS1-VS2. Item has been tested and guaranteed to be solid PLATINUM. This ring is in excellent pre-owned condition. Please disregard the two characters at the end of the title. You are using for inventory purposes. As a courtesy, please notify us of any return. Always fast & free unless otherwise stated. Collectors Coins & Jewelry has been family owned and operated on Long Island, NY since 1946. We have four brick and mortar locations and offer the highest quality products with unbeatable customer service. The item “1920s Antique Art Deco Solid Platinum. 90ctw Old Mine Cut Diamond Navette Ring” is in sale since Thursday, November 30, 2017. This item is in the category “Jewelry & Watches\Jewelry\Gold & Watches & Brooches\Diamond & Gemstones”. The seller is “collectorsbuysell” and is located in Huntington, New York. This item can be shipped worldwide.
The island is hilly, rising to 119 m at Mt. Wanim. The island is 0.9 km² in area, it is part of the Pana Tinani Group. The island is 0.9 km south of Pana Tinani, and separated from it with the Bulami Channel. The island was discovered in the late 18th century. At the census of population in 2014, the island had 600 inhabitants, spread across 3 small villages. The main town is Bunbun, located on the northwest point.

At the census of population in 2014, the island had 600 inhabitants, spread across 3 small villages. The main town is Bunbun, located on the northwest point.

The submarine was laid down in Rotterdam at the shipyard of Rotterdamsche Droogdok Maatschappij on 31 May 1930. The launch took place on 11 July 1931. On 6 July 1933 the boat was commissioned in the Dutch navy. On 7 February 1934 K XIV and K XV left the Netherlands for the Dutch East Indies. The route they took led through the Suez Canal. On 6 September 1938 they participated in a fleet show at Surabaya. The show was held in honor of the Dutch Queen Wilhelmina of the Netherlands who celebrating her 40th year as head of state. More than twenty navy ships participated in the show. In the war K XIV sank several Japanese ships. She survived the war and was decommissioned on 23 April 1946. 1 June 1946 she was stricken and sold for scrap in December 1950.

MISSOURI VALLEY CONFERENCE Men's Soccer Tournament is the culmination of the regular season. The regular season conference matches determine the seeding in the tournament, which determines the conference's automatic berth into the NCAA Tournament. All teams in the Missouri Valley Conference, or MVC, play each other once during the season. Teams play certain teams at home during even number years, and then will play those teams on the road during odd number years. Teams are awarded three points for a win, a point for a draw and no points for a loss. In the event that teams are tied on points, the first tiebreaker is head-to-head record. If that tiebreaker is tied, goal differential is applied, followed by goals scored, then away goals, then RPI. Missouri State won the regular season with a 5-2-1 record.

The 2017 Missouri Valley Conference Men's Soccer Tournament is the culmination of the regular season. The regular season conference matches determine the seeding in the tournament, which determines the conference's automatic berth into the NCAA Tournament. All teams in the Missouri Valley Conference, or MVC, play each other once during the season. Teams play certain teams at home during even number years, and then will play those teams on the road during odd number years. Teams are awarded three points for a win, a point for a draw and no points for a loss. In the event that teams are tied on points, the first tiebreaker is head-to-head record. If that tiebreaker is tied, goal differential is applied, followed by goals scored, then away goals, then RPI. Missouri State won the regular season with a 5-2-1 record.

...
Rodrigo Tarín, born in Chiva, Valencia, Community, Tarín joined FC Barcelona’s youth categories in 2011, from Valencia CF. On 18 September 2014, he renewed his contract until 2018, and was promoted to the reserves in Segunda División B the following July. On 22 August 2015, starting in a 1–2 away loss against UC Cornellà. He scored his first senior goal on 17 September of the following year, netting the winner in a 2–1 home success over CD Atlético Baleares; in November, however, he suffered a knee injury which took him out for six months.

On 19 August 2017, starting in a 2–1 away win against Real Valladolid for the Segunda División championship. The following 27 June, he signed a three-year deal with La Liga side CD Leganés. He made his debut in the main category of Spanish football on 26 September 2018, starting in a 2–1 home defeat of former side Barcelona. On 17 September of the following year, netting the winner in a 2–1 home success over CD Atlético Baleares; in November, however, he suffered a knee injury which took him out for six months.

On 19 August 2017, starting in a 2–1 away win against Real Valladolid for the Segunda División championship. The following 27 June, he signed a three-year deal with La Liga side CD Leganés. He made his debut in the main category of Spanish football on 26 September 2018, starting in a 2–1 home defeat of former side Barcelona.

Alexis Gutiérrez, born in Chiva, Valencia, Community, Gutiérrez at a young age was scouted and joined Guadalajara’s youth academy in 2012. He then continued through Chivas Youth Academy successfully going through U-13, U-15, U-17 and U-20. Until finally receiving attention to join Cruz Azul, Pedro Caixinha being the coach promoting Gutiérrez to the first team. On 18 September 2014, he renewed his contract until 2018, and was promoted to the reserves in Segunda División B the following July. He then continued through Cruz Azul as Youth Academy successfully going through U-13, U-15, U-17 and U-20. Until finally receiving attention to join Cruz Azul as the first team. Pedro Caixinha, former Pumas being the coach promoting Gutiérrez to the first team. On 18 September 2014, he renewed his contract until 2018, and was promoted to the reserves in Segunda División B the following July.

On 22 August 2017, starting in a 2–1 away win against Real Valladolid for the Segunda División championship. The following 27 June, he signed a three-year deal with La Liga side CD Leganés. On 18 September 2014, he renewed his contract until 2018, and was promoted to the reserves in Segunda División B the following July. On 22 August 2017, starting in a 2–1 away win against Real Valladolid for the Segunda División championship. The following 27 June, he signed a three-year deal with La Liga side CD Leganés.

Figure 30: Validation / training example pair from Wiki40B:en with intermediate to low influence. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
Table:

<table>
<thead>
<tr>
<th>Index</th>
<th>Estim. Link and Text</th>
<th>infl</th>
<th>0.0560</th>
</tr>
</thead>
<tbody>
<tr>
<td>train</td>
<td>2079293</td>
<td>men</td>
<td>0.2117</td>
</tr>
<tr>
<td><a href="https://www.capitalfm.com/northeast/terms-conditions/i-saw-it-first-1000-april19/">https://www.capitalfm.com/northeast/terms-conditions/i-saw-it-first-1000-april19/</a></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- I Saw It First Radio competition (the “Competition”) run from Monday 8th April 2019 to Saturday 27th April 2019.
- Online entry will open at 00:01 on Monday 8th April 2019 and will close at 17:30 on Saturday 27th April 2019.
- The winner will be picked at random from all correct entries and notified via telephone or email. Eligibility: Entrants must be 18 or over. Entry is restricted to one entry per person; duplicate entries will be excluded from the Competition. Prize: 1. One winner will receive the £1000 prize package which consists of: A £600 I Saw It First voucher (valid for 3 months) worth £600 (the “Voucher”).

---

Figure 31: Generated / training example pair from RealNews with high to intermediate influence. The generated examples are directly taken from publicly released generations of the Grover-Mega (p=0.96) model [Zellers et al., 2019]. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python difflib.
The 930-kilometer long TAP project envisages transportation of gas from Azerbaijan’s Shah Deniz Stage 2 to the EU countries. The pipeline will connect to the Trans Anatolian Natural Gas Pipeline (TANAP) on the Turkish-Greek border, run through Greece, Albania and the Adriatic Sea, before coming ashore in Italy’s south. TAP’s shareholding is comprised of BP (20 percent), SOCAR (20 percent), Snam S.p.A. (20 percent), Fluxys (19 percent), Enagás (16 percent) and Axpo (5 percent). — Follow the author on Twitter: @Ali_Mustafayev Follow Trend on Telegram. Only most interesting and important news.

Baku, Azerbaijan, April 15 By Ali Mustafayev – Trend: Italy would have to import almost 50 percent of its natural gas through Germany without the Trans Adriatic Pipeline (TAP), managing director of TAP AG Beat Rathmann told Trend April 15. TAP is a part of the Southern Gas Corridor project, aimed at diversifying the European energy supply sources and routes. The 930-kilometer long TAP project envisages transportation of gas from Azerbaijan’s Shah Deniz Stage 2 to the EU countries. The pipeline will connect to the Trans Anatolian Natural Gas Pipeline (TANAP) on the Turkish-Greek border, run through Greece, Albania and the Adriatic Sea, before coming ashore in Italy’s south. TAP’s shareholding is comprised of BP (20 percent), SOCAR (20 percent), Snam S.p.A. (20 percent), Fluxys (19 percent), Enagás (16 percent) and Axpo (5 percent). — Follow the author on Twitter: @Ali_Mustafayev Follow Trend on Telegram. Only most interesting and important news.

Figure 32: Generated / training example pair from RealNews with intermediate to low influence. The generated examples are directly taken from publicly released generations of the Grover-Mega (p=0.96) model [Zellers et al., 2019]. Red / green highlighted text indicate deleted / added text in the training example comparing to the corresponding validation example, generated using Python diffLib.
NEW DELHI: India is likely to see average monsoon rains this year, the state-run weather office said on Monday, which should support agricultural production and economic growth in Asia’s third-biggest economy, where half of the farmland lacks irrigation. Monsoon rainfall is expected to be 96 percent of the long-term average, M. Rajeevan, secretary at the Ministry of Earth Sciences, told a news conference. The India Meteorological Department (IMD) defines average, or normal, rainfall as between 90 percent and 104 percent of a 50-year average of 89 centimeters for the entire four-month season beginning June. “Overall, the country is expected to have well distributed rainfall scenarios during the 2019 monsoon season, which will be beneficial to farmers in the country during the ensuing kharif (summer sowing) season,” the IMD said in its forecast. India’s only private weather forecasting agency, earlier this month forecast rainfall could be below normal this year. Monsoon rains, the lifeline for India’s farm-dependent $2.8 trillion economy, arrive on the southern tip of Kerala state around June 1 and retreat from the desert state of Rajasthan by September. After a wet spell, sowing of summer season crops gets off to a strong start, boosting crop yields and output which is in turn raises rural incomes and usually lifts consumer spending in India. Plentiful monsoon rains lift agricultural production this year, which could keep food prices under control. Subdued overall inflation could also add to pressure on India’s central bank to cut interest rates. “We’re a fair forecast, showing near-normal and well-distributed rainfall, will bode well for near-term food inflation,” said Madhulika Arora, lead economist at Edelweiss Securities, Fix and Rates. The next policy review by India’s central bank is scheduled for June 4, after the country’s election. Millions of Indians are casting their votes in a mammoth general election, spread over seven weeks. On the downside, higher production could mean farmers continue to get hit by crop prices, a major cause for concern in rural India, where most Indians live, in the past two years. After falling for five straight months, retail food prices in India rose 0.30 percent in March from a year earlier. Last month, a senior IMD official told Reuters that this year’s monsoon was likely to be robust and healthy provided there wasn’t a surprise El Nino phenomenon. “El Nino is weakening and we expect that El Nino will get weakened further. There is no reason to be worried about El Nino,” Rajeevan said. A strong El Nino, marked by a warming of the sea surface on the Pacific Ocean, can cause severe drought in Australia, Southeast Asia and India, while disrupting other parts of the world such as the US Midwest and Brazil. Rains are expected in five quarters, compared with four last year and also the five-year average. The IMD has forecast an overall 6.6 percent in the December quarter, from 7.0 percent in the previous period and the slowest in five years over the past two decades, even after taking into account an error band of plus or minus 5 percentage points.

According to the Better For Global Markets (p=0.96) model [Zellers et al., 2019], the generated examples are directly taken from publicly released generations of the Grover-Mega model. Red text indicates deleted text, green text indicates added text, and blue text indicates unchanged text in the training example.