A Fast and Effective Solution to the Problem of Look-ahead Bias in LLMs

113

Humzah Merchant, Bradford Levy

TL;DR: A key barrier to the application of LLMs in finance is look-ahead bias. Using two small models to guide the generation of a larger model can reduce this to allow effective backtesting of LLMs.

Finance Problem: Look-ahead Bias

LLMs have potential for trading and fundamental research. However, backtesting within their training knowledge cutoffs can lead to analysis based on information leaked from the future.

```
System: Today is Dec 31, 2019.
User: What stocks should I add to my
portfolio?
Assistant: Buy Zoom, short airlines
```

Unfortunately, training a frontier model with an earlier knowledge cutoff is extremely expensive.

Method: Divergence Decoding

We can approximate a model being trained to an earlier knowledge cutoff by adjusting its logits with two smaller models: one trained on data before the cutoff, and one trained on all data.

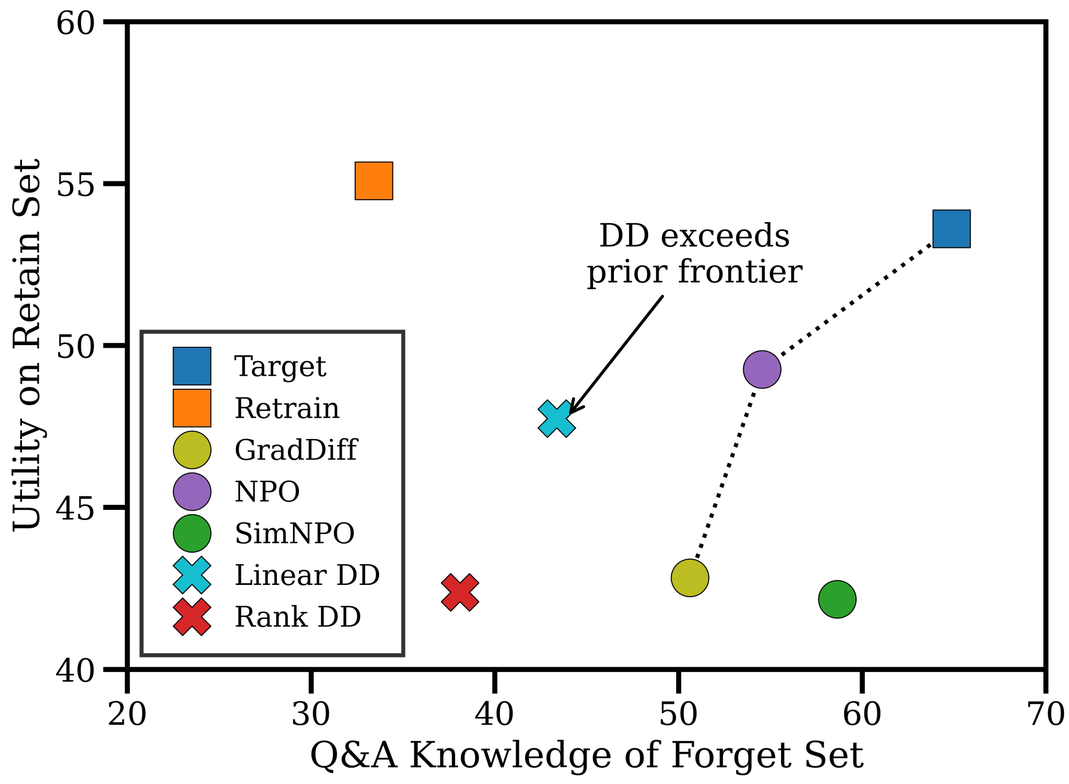
$$l_Q^{LC}(x_{<t}) = l_P(x_{<t}) + \alpha \cdot \left[l_q(x_{<t}) - l_p(x_{<t})\right]$$

$$l_Q^{R}(x_{<t}) = l_P(x_{<t}) - \mathbb{1}_{rank\left(l_p(x_{<t}) - l_q(x_{<t})\right) \leq k} \cdot \infty$$

Since the total cost to train a model is approximately $O(n^2)$, this is significantly cheaper than retraining the large model.

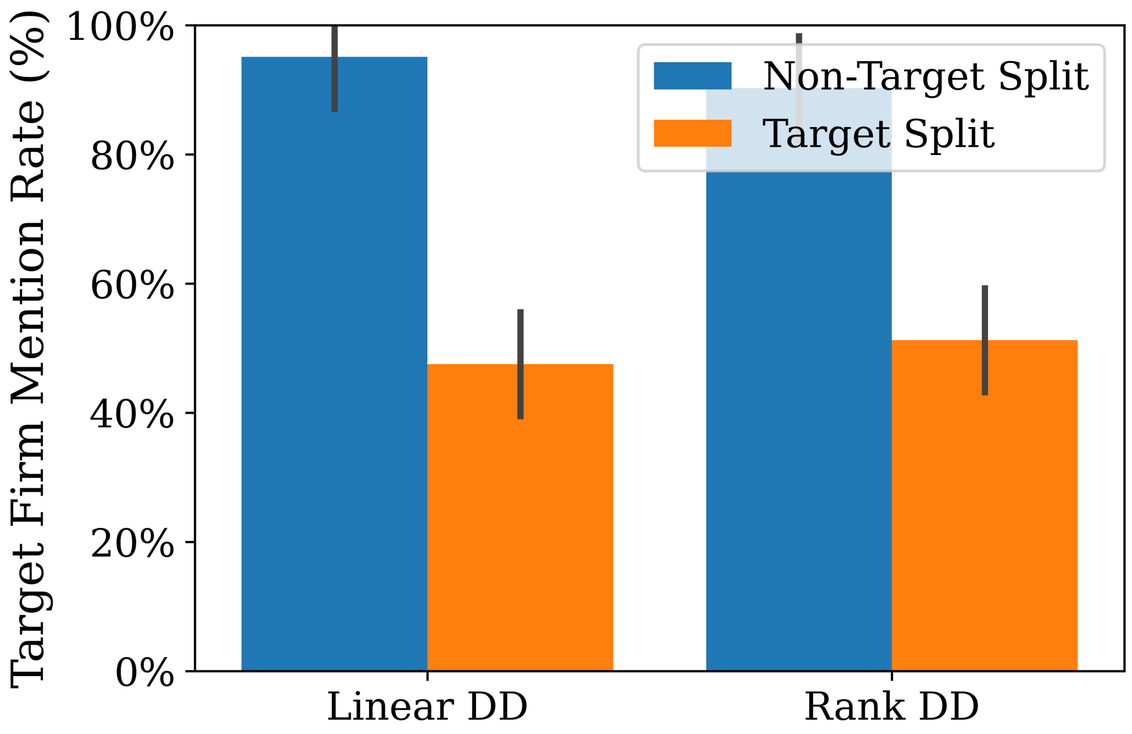
General Problem: Unlearning

Look-ahead bias arises due to semantic and exact memorization of training data. In a similar way, AI safety researchers explore similar problems in copyrighted content regurgitation and toxic content generation. Methods to remove the information from outputs without retraining from scratch are referred to as **unlearning** methods.



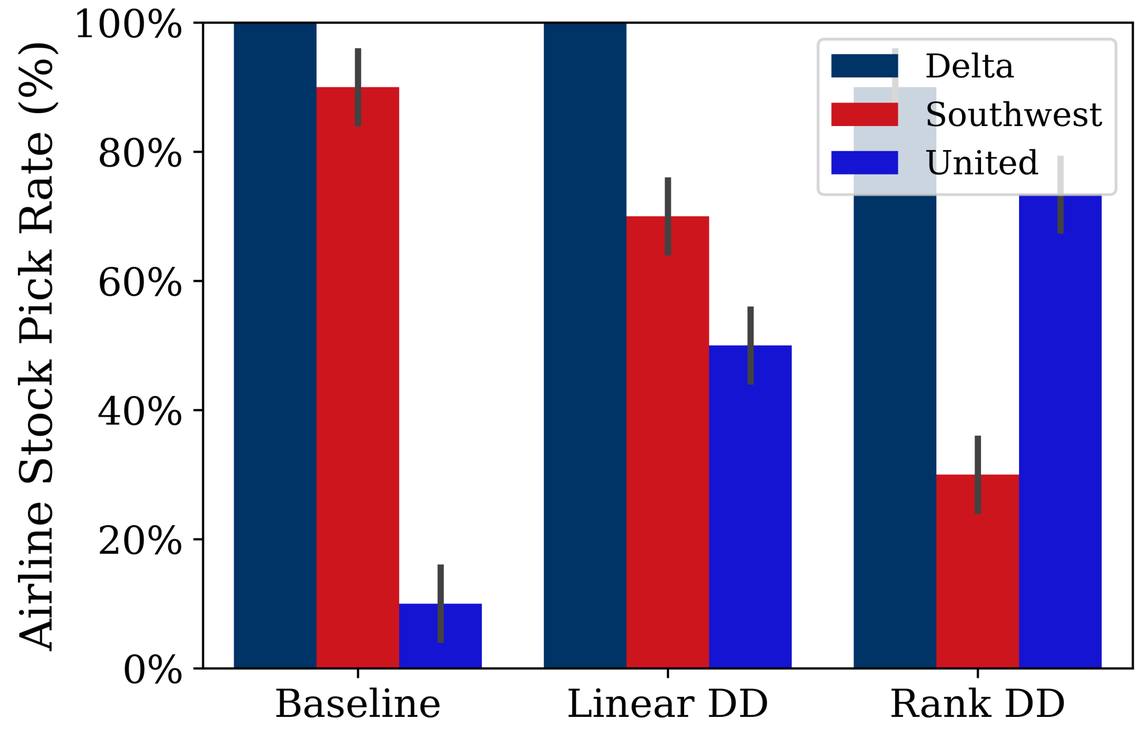
We test on MUSE, a key unlearning benchmark. Most unlearning methods require weight updates on the large model, making them expensive and often causing large utility loss – a phenomenon known as catastrophic forgetting.

Predicting M&A Transactions



We sampled M&A deals that Gemma 3 27B had memorized and split them in two sets. We found the ability to effectively unlearn the target split while retaining utility on the non-target split.

Stock Sentiment



When asked to assemble a diverse portfolio, Gemma 3 27B has a strong preference towards Delta and Southwest due to historical sentiment. If we finetune p on older (2014-2016) reports of the airlines' performances and q on recent (2022-2024) reports, we can change the portfolio construction choices. *Of course, for look-ahead bias, we would do the opposite.*

CHICAGO BOOTH
THE UNIVERSITY OF CHICAGO BOOTH SCHOOL OF BUSINESS
Center for Applied Artificial Intelligence