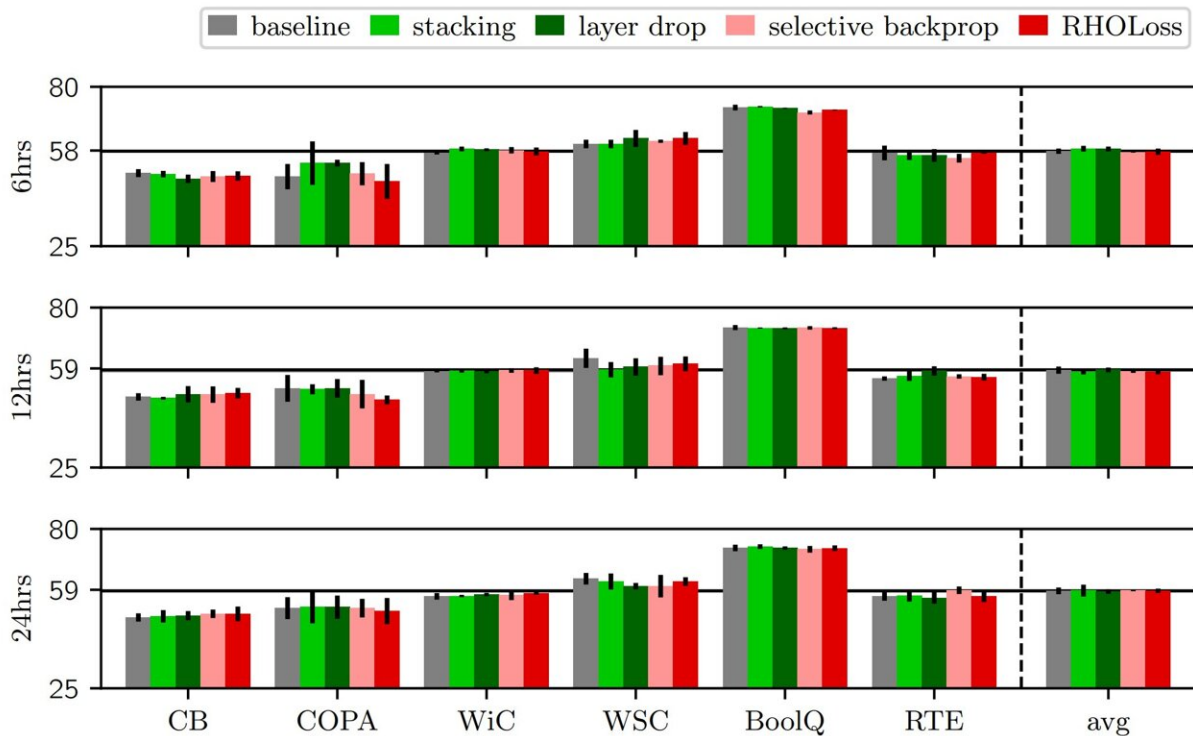


# Machine-training workarounds found to yield little advantage

August 7 2023, by Peter Grad



BERT models evaluated on SuperGLUE. The black vertical error bars indicate the standard deviation over three seeds. The black horizontal line shows the baseline average performance. Credit: *arXiv* (2023). DOI: [10.48550/arxiv.2307.06440](https://doi.org/10.48550/arxiv.2307.06440)

An AI-infused future has tremendous promise in virtually all aspects of our lives from medicine to education, industry to finance.

But it comes with a cost—literally. To train GPT-4, OpenAI had to shell out \$100 million, according to the company's CEO Sam Altman.

Although some chatbots are offered free of charge, [small businesses](#) seeking to create a chatbot service would pay about \$4,000, according to an Accubits blog. And according to AI company Figure Eight, creating a dataset for machine learning can cost up to \$100 per task. Furthermore, a Stanford University study found that labeling just a [single image](#) dataset for machine learning costs \$3.50 per image. Rentals of required cloud computing platforms can top \$5 an hour.

So economizing is a concern among potential users.

Researchers at the University College London and University of Edinburgh designed a measurement system that compared results from standard machine learning approaches and alternative means eyed by companies as cost-saving measures. They found alternate machine training approaches yielded only minimal results.

They conducted research on several variations of three main categories of efficiency models.

"In most cases, these methods—which are often quite a bit more complicated and require more implementation efforts—in our experiments didn't really result in a significant improvement," said Oscar Key, of the University College London and a co-author of the report.

Three categories the team looked at were:

Batch selection, which involves processing groups of data bits rather than the individual components. A simple example would be appending a word to the filename of a large number of digital photos to more clearly identify them. A batch operation that renames them in an instant is

quicker, and ultimately cheaper, than altering the names one by one.

Layer stacking, which employs multiple layers of neural network units while sequentially processing data. It permits models to learn sophisticated language patterns and relationships. For example, models use stacking to recognize grammar, semantics and sentiment in language input and can shape output accordingly in real time. Style, tone and voice are modified according to prompts on various neural network layers.

Efficient optimizers, which as their name implies, are algorithms that aim to accelerate search functions, minimize wasteful operations, hasten the [learning process](#) and ideally obtain better solutions. The researchers used the relatively new Sophia optimizer, which is said to be twice as fast as the more commonly used, state-of-the-art Adam optimizer.

Despite successfully skipping irrelevant data, ignoring less [relevant data](#) and optimizing helpful and pertinent data, these approaches resulted in inferior output.

Layer stacking was the only process that resulted in "training and validation" gains, though they were minimal, according to the report. But such gains "vanish" the longer the training was performed.

In other words, optimization of [machine learning](#) may use less computing power and be less expensive, but results are inferior, and can be improved only through additional [training](#) and cost.

"Training models to get to even reasonable performances is usually very expensive," says Jean Kaddour, another author of the report.

The report, "No Train No Gain: Revisiting Efficient Training Algorithms For Transformer-based Language Models," is published on the *arXiv* preprint server.

**More information:** Jean Kaddour et al, No Train No Gain: Revisiting Efficient Training Algorithms For Transformer-based Language Models, *arXiv* (2023). [DOI: 10.48550/arxiv.2307.06440](https://doi.org/10.48550/arxiv.2307.06440)

© 2023 Science X Network

Citation: Machine-training workarounds found to yield little advantage (2023, August 7) retrieved 11 May 2024 from <https://techxplore.com/news/2023-08-machine-training-workarounds-yield-advantage.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.