**Write a brief review of the story. Your review should cover the main idea and give your reaction to it, focusing on possible applications.**

<https://ai.googleblog.com/2017/04/federated-learning-collaborative.html>

One of the hottest fields today is machine learning and big data. Both ideas are centered on the idea that by aggregating data millions of user’s trends, we can best predict what to show to future users. Developments in the fields of distributed systems and cloud infrastructure have made it easier to centralize the training data on one machine or datacenter, but this has a few downsides. One of which is that a user’s data is being harvested and sent to the cloud, where it is stored, and therefore vulnerable to data breaches. Another disadvantage is the latency of updates in the model, having to do with the time to send new training data to the cloud, recompute the model, and then redeploy the model back to the user’s device.

Lastly, consider the affect of a single drop in a large bucket of water, analogous to one’s data being fed through a model that has already been trained on millions of datapoints. The model is not going to change much because of one user, since that

Federated Learning is a new approach that attempts to tackle some of these challenges by collaboratively learning a shared model, while keeping an individual’s data on the device, meaning massive centralized training sets are no longer needed in the cloud.

Federated Learning works as follows:

1. Download the current master model to the device
2. As your device produces more data, train the locally stored model
3. Batch up the changes and send the update to the cloud (via encrypted channels)
4. In the cloud, process the update and make chances to the master model

All the training data remains on your device, and no individual updates are stored in the cloud.

This idea has been around for a while, but the recent news is that a team of engineers at Google have figured out how to deploy this at scale on the GBoard product (android keyboard).

When I heard about this new technology, I was surprised that there was a way to train models without centralized datasets. In my (limited) experience with machine learning models, I’ve always figured that the data had to be in the same place. I think this technology can be really disruptive, especially in light of recent privacy concerns. To me, the most important benefit is that my data is not being sent to Google or Facebook or whatever app I am using, but rather only an encrypted ‘changelog’ of updated model weights.

The applications of this technology are immense. The authors of the paper concede that certain machine learning problems, such as recognizing dog breeds, cannot be solved using Federated Learning, but some domains include improving the language models based on what you actually type on your phone (since people’s phones can have different layouts) and photo rankings based on what kinds of photos people look at, share, or delete. Another realm of potential impact is in advertising – where there is currently a lot of buzz regarding data harvesting and selling. A common critique is that “Facebook is selling my data to marketers”, and while that statement may or may not be true, it’s unquestionable that people would be more comfortable if their personal data was not directly handled by the platform that they are using.