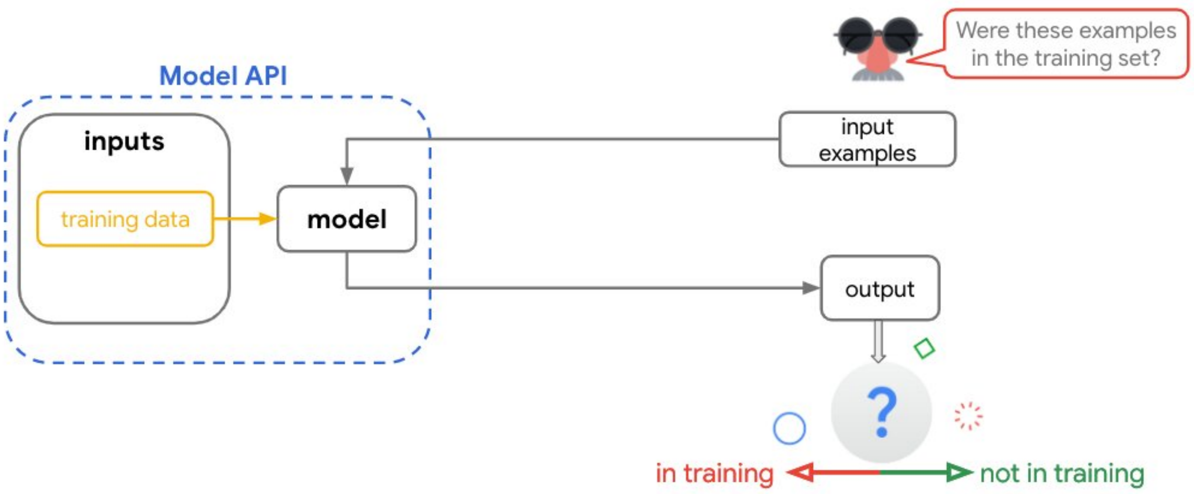
Google releases TensorFlow privacy testing module

June 25 2020, by Peter Grad



Credit: Google

Google released a toolset accessory this week that will allow developers working with machine learning models to better rein in leaks of private data.

The TensorFlow Privacy module applies a newer means of testing for vulnerabilities in massive datasets containing information used for such purposes as medical care, facial recognition and surveillance.

Google established the TensorFlow Privacy library a year ago to help developers achieve greater accuracy in machine learning projects while reducing risk of compromising personal data of subjects contained in databases. Google stated at the time, "Modern machine learning is increasingly applied to create amazing new technologies and user experiences, many of which involve training machines to learn responsibly from sensitive data, such as personal photos or email. We intend for TensorFlow Privacy to develop into a hub of best-of-breed techniques for training machine-learning models with strong privacy guarantees."

At its 2019 launch, TensorFlow Privacy relied on a concept known as differential privacy, in which patterns of groups within datasets are shared in public while links to individuals comprising the datasets are shielded. In deep learning applications, developers generally aim to encode generalized patterns rather than specific details that help identify participants and threaten anonymity.

One means of accomplishing this is by introducing limited 'noise' that helps protect identities of users. Such noise, however, carries the risk of degrading accuracy.

Google took a new approach with the TensorFlow module announced this week. Applying a test called 'membership inference attack,' TensorFlow Privacy can establish a score revealing how vulnerable a model is to leaked information.

"Cost-efficient membership inference attacks predict whether a specific piece of data was used during training," Google stated in its TensorFlow blog Wednesday. "If an attacker is able to make a prediction with high accuracy, they will likely succeed in figuring out if a data piece was used in the training set. The biggest advantage of a membership inference attack is that it is easy to perform and does not require any re-training."

"Ultimately," Google said, "these tests can help the developer community identify more architectures that incorporate privacy design principles and data processing choices."

Google sees this as a starting point for "a robust privacy testing suite" that, because of its ease of use, can be used by machine learning developers of all skill levels.

As more institutions rely on massive data-hungry machine learning projects, privacy concerns are heightened. Last year Microsoft was forced to remove more than 10 million images from its globally distributed MS Celeb facial recognition training program after learning subjects were not asked for their permission before publication. Google hastily abandoned a health data-share project with Ascension following growing concerns that chest X-rays could expose personal information.

And Apple and Google have drawn criticism over weak privacy protections surrounding the use by tens of millions of users of AI agents such as Siri and Google Assistant in which audio recordings from people's phones and inside their homes were reportedly stored and accessed without authorization.

More information: blog.tensorflow.org/2020/06/in … testing-library.html

github.com/tensorflow/privacy

Citation: Google releases TensorFlow privacy testing module (2020, June 25) retrieved 9 April 2024 from https://techxplore.com/news/2020-06-google-tensorflow-privacy-module.html