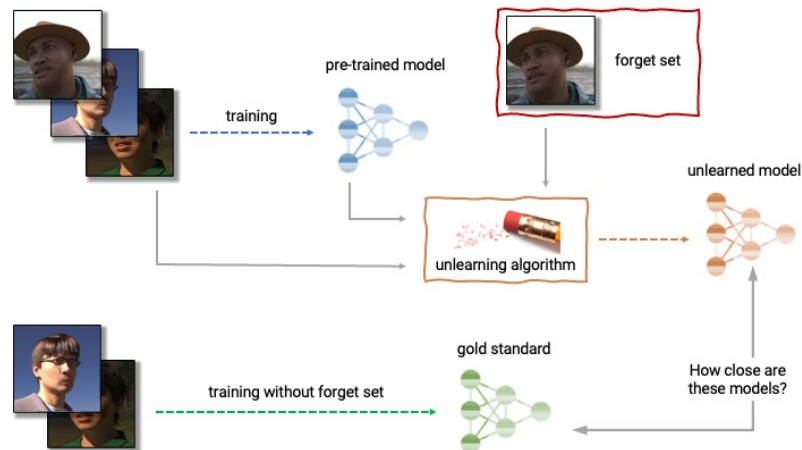# Machine Unlearning Challenge

Deep learning has revolutionized applications like image generation and language models, but caution is needed due to biases and privacy concerns. Deleting data completely is challenging, as it involves removing its influence on trained models. Recent research shows membership inference attacks can determine if data was used for training, even after deletion.

Machine unlearning aims to remove the influence of specific training examples from a model. An ideal algorithm maintains accuracy while removing the "forget set." Retraining can be costly, so an efficient method that modifies the existing model is desired.

Research scientists at Google Deepmind have announced the first Machine Unlearning Challenge (see here and here). The challenge is centered on a scenario in which an age predictor is built from face image data and, after training, a certain number of images must be forgotten to protect the privacy or rights of the individuals concerned.

In this project, you will come up with new ways to untrain deep neural networks and participate in the challenge.

Requirements: Strong motivation, knowledge in deep learning, or a solid background in machine learning. Previous experience with Python and libraries such as TensorFlow or PyTorch is an advantage. Looking into the competition and the referenced starting kit before applying is recommended. We will have weekly meetings to discuss open questions and determine the next steps.

Interested? Please contact us for more details!

Contact

- Luca Lanzendörfer : `lanzendoerfer@ethz.ch`, ETZ G93

- Benjamin Estermann: `besterma@ethz.ch`, ETZ G60.1