

CLAIMS

1. A computer-implemented method for private transfer learning, comprising:
generating a machine learning model comprising a training application programming interface (API) and an inferencing API;
encrypting the machine learning model using a predetermined encryption mechanism;
copying the encrypted machine learning model to a trusted execution environment; and
executing the machine learning model in the trusted execution environment using at least one of the training API and inferencing API.
2. The method of claim 1, wherein the machine learning model comprises a deep neural network (DNN) model.
3. The method of claim 2, wherein the DNN model is trained to perform a generic task.
4. The method of claim 3, wherein the training API trains the DNN model to perform a task that refines the generic task to a more specific task than the generic task.
5. The method of any of the preceding claims, further comprising generating a combination of a body of the machine learning model with a head of an additional machine learning model in the trusted execution environment.
6. The method of claim 5, further comprising performing the inferencing API for the combination.
7. The method of either of claims 5 or 6, further comprising performing the training API for the combination.
8. A system for private transfer learning, the system comprising:
a computer processing circuit; and
a computer-readable storage medium storing instructions, which, when executed by the computer processing circuit, are configured to cause the computer processing circuit to perform a method comprising:
generating a machine learning model comprising a training application programming interface (API) and an inferencing API;
encrypting the machine learning model using a predetermined encryption mechanism;
copying the encrypted machine learning model to a trusted execution environment;
executing the machine learning model in the trusted execution environment using the training API; and
executing the machine learning model trained by the training API by using the inferencing API.

9. The system of claim 8, wherein the machine learning model comprises a deep neural network (DNN) model.
10. The system of claim 9, wherein the DNN model is trained to perform a generic task.
11. The system of claim 10, wherein the training API trains the DNN model to perform a task that refines the generic task to a more specific task than the generic task.
12. The system of any of claims 8 to 11, the method further comprising generating a combination of a body of the machine learning model with a head of an additional machine learning model in the trusted execution environment.
13. The system of claim 12, the method further comprising performing the inferencing API for the combination.
14. The system of either of claims 12 or 13, the method further comprising performing the training API for the combination.
15. The system of any of claims 8 to 14, wherein:
 - executing the machine learning model in the trusted execution environment using the training API comprises executing the training API on the GPU; and
 - executing the machine learning model trained by the training API by using the inferencing API comprises executing the inferencing API on the GPU.