Data Use Agreement for Open AI Model Development (DUA-OAI)

Goal

Sharing data can help address some of society's biggest challenges and help individuals and organizations be more innovative, efficient, and productive. We want to make it easier for individuals and organizations that want to share data to do so. We're working with companies, academics, and researchers to build better processes and tools. As a first step, we've taken a closer look at a specific data use scenario with this Data Use Agreement for Open AI Model Development (DUA-OAI). The goal of the DUA-OAI is to create a template agreement that parties might use to share data to train an artificial intelligence (AI) model and then to make that trained AI model publicly available through an open source licensing structure.

For more information on Microsoft's resources to Removing Barriers to Data Innovation, visit here.

Overview

This template is designed to allow a data holder to share data for the limited purpose of training an AI model. While this agreement contemplates that the AI code and the trained AI model associated with it be made freely available under an open source licensing model, the data used to train the AI model is shared only for the limited purpose of training that AI model and otherwise remains confidential.

Contemplated use case

We envision that this agreement is suitable for situations where the parties want to share data to train an AI model when that data cannot be made public because of privacy and/or business concerns.

This agreement contemplates that the parties might use various contractual terms or technical mechanisms to address the sensitivities of the data shared. For example, and perhaps most simply, if the use of the data to train the AI model does not implicate the privacy or business sensitivity of the data, then the limitation of use of the data in the agreement may be sufficient to address those concerns. In other situations, the parties might use other techniques to address these concerns including data aggregation, data pseudonymization, or limiting or prohibiting human inspection of some or all of the data. In more complex scenarios, the parties might use technological privacy enforcement practices including differential privacy, data brokering, or requiring data be processed inside trusted security spaces such as via containers, data bricks, cryptographic kernels, or with the assistance of a trusted third party.

With this agreement, Microsoft is not giving legal advice. Please consider your own circumstances and seek your own legal counsel as needed.

Contributing

This DUA-OAI template might be improved as we learn more from users, and it does not address every data sharing scenario. We are interested in your feedback both as to how the DUA-OAI might be made more useful and as to other templates that address other common data sharing scenarios. Please submit your comments to datainno@microsoft.com.

In this process, we hope to listen to your feedback, learn from it, and improve and expand this work product over time.