

# Computational Use of Data Agreement (C-UDA)

## Goal

Sharing data can help address some of society's biggest challenges and can 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 Computational Use of Data Agreement (C-UDA), intended to complement the Open Use of Data Agreement (O-UDA). The goal of the C-UDA is to define a use of data sets for AI training purposes that contain third party materials, in a manner consistent with law. We hope to gather community input that evolves the agreement for broad use. Our aim is to release a v1 of the C-UDA in Fall 2019. Please provide feedback by October 1, 2019.

For more information on Microsoft's resources to Removing Barriers to Data Innovation, visit [here](#).

## Overview

The C-UDA is a simple agreement that allows the data holder to make data available to anyone for computational use purposes, such as artificial intelligence, machine learning, and text and data mining. In short:

- It is intended for data sets that may include material not owned or controlled by the data distributor.
- It addresses data that is assembled from lawfully accessed, publicly available sources to be used for computational analysis.
- Redistribution of the Output from use of the data under the agreement—including results of analysis of the data or ML models trained with the data—carries no obligations.
- Redistribution of data under the agreement—modified or unmodified—requires use of the C-UDA.
- The redistribution obligations are designed to encourage sharing by limiting the liability of the data provider and ensuring that those downstream can identify where the data came from.

## Contemplated use case

We envision that this agreement is suitable for situations where the original data provider owns or has lawfully acquired the material in the data set (because they have express permission to use the material), or where they have assembled materials from lawfully and publicly accessible sources and the data is appropriate for distribution for computational use purposes. Permission to redistribute this material is limited to computational analysis to remain compliant with legal precedent and statutory exceptions and to respect the legitimate interests of

third party rights owners.

This agreement is not recommended where the data provider includes material in the data set that (i) was not lawfully accessed and is not appropriate for distribution for computational use purposes, (ii) is subject to a legally binding restriction that restricts its further distribution, or (iii) raises privacy concerns arising from its distribution. Data Providers may need to consider whether additional measures are appropriate to ensure that data is not made available for use beyond legally permissible computational uses.

A limitation of this agreement is that it does not authorize uses beyond computational use that may otherwise be legally permissible.

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

## **The C-UDA does not meet the Open Data Definition**

The C-UDA is not intended to be and should not be described as an open data license. Specifically, it does not permit use for any purpose as described in Section 2.1.1 and 2.1.8 of the Open Definition. The C-UDA is intended to address situations in which data cannot be shared under an open license, but it is possible for a data provider to permit computational use. For situations in which an open data license is appropriate, see the O-UDA and other open data licenses.

## **Why a “computation” only agreement?**

We developed the C-UDA to address a gap among current public agreements. Data that is useful for computational analysis may often include copyrightable content, and global legal precedent and legislation have confirmed that copyrighted works may be used for computational use without express consent of the owner. However, continued perceived uncertainty over copyright law has caused many data providers to resort to limitations that significantly restrict who can use data, or how the data can be used, in ways that may be more restrictive than those permitted by applicable law or legislation. These restrictions may create uncertainty or cause confusion among users that greatly limits the usefulness and benefit of data sets containing copyrighted works in artificial intelligence activities, such as machine learning. The C-UDA does not restrict who can use such data, but it limits the use of data to computational analysis to be consistent with applicable law and legislation, and to respect the legitimate interest of rights holders.

## **Contributing**

This project welcomes contributions and suggestions under CC0-1.0. To suggest edits, open a Pull Request or to start a discussion open an Issue. Or, if you prefer

to submit comments via email, please submit them to [datainno@microsoft.com](mailto:datainno@microsoft.com). If you wish your comments to remain anonymous, please submit them by email and say so in the first line of the email.

This project has adopted the Microsoft Open Source Code of Conduct. For more information see the Code of Conduct FAQ or contact [opencode@microsoft.com](mailto:opencode@microsoft.com) with any additional questions or comments.

## **Legal Notices**

Microsoft and any contributors grant you a license to content in this repository under CC0-1.0, see the LICENSE file.