

Computational Use of Data Agreement v0.1

Annotated-Discussion-DRAFT-20190722

This is the Computational Use of Data Agreement, Version 0.1 (the “C-UDA”). Capitalized terms are defined in Section 5. Data Provider and you agree as follows:

Comment *The C-UDA was developed for use by a Data Provider that owns or controls Data, or has assembled Data from lawfully accessed, publicly available sources, and wishes to limit the use for computational purposes to be consistent with copyright laws. The C-UDA is not intended for a use of Data that may include personal data. To be precise, it is not appropriate for data sets that include any data that might include materials subject to privacy laws such as the GDPR or HIPAA.*

1. Provision of the Data

1.1. You may use, modify, and distribute the Data made available to you by the Data Provider under this C-UDA for Computational Use if you follow the C-UDA’s terms.

Comment: *The C-UDA permits data to be used for computational use only, but it also allows the Data to be modified and redistributed so long as the Downstream Recipient also complies with the C-UDA’s terms. Because the rights (whether copyright, database rights, or merely access rights) that may potentially apply to content included in a data set can vary around the world, the C-UDA is styled to permit a set of uses recognized under law rather than to grant specific rights that may or may not be applicable.*

1.2. Data Provider will not sue you or any Downstream Recipient for any claim arising out of the use, modification, or distribution of the Data provided you meet the terms of the C-UDA.

Comment: *This is a promise by the Data Provider not to sue the user so long as the user complies with C-UDA’s requirements. It doesn’t allow a Data Provider to terminate a permitted use of the Data, but it does allow the Data Provider to bring an action to enforce the C-UDA’s terms.*

2. Restrictions

2.1 You agree that you will use the Data solely for Computational Use.

Comment: *The agreement allows data provided under this C-UDA to be used for computational use (e.g., to train AI models), while not permitting other uses that might conflict with rights that may be held by third parties in the material within a database, such as broad rights to copy and distribute expressive works.*

2.2 The C-UDA does not impose any restriction with respect to the use, modification, or distribution of Outputs.

3. Redistribution of Data

3.1. You may redistribute the Data, so long as:

3.1.1. You include with any Data you redistribute all credit or attribution information that you received with the Data, and your terms require any Downstream Recipient to do the same; and

3.1.2. You bind each recipient to whom you redistribute the Data to the terms of the C-UDA.

Comment: *The only requirements for redistributing Data are to maintain attribution (if any) and use restrictions under the C-UDA, so that Downstream Recipients are bound by it for their use. These two restrictions apply only to the Data, but not to Output. Maintaining attribution is an accepted practice for sharing data to indicate its source or provenance. Requiring the use of the C-UDA for subsequent distribution provides further certainty to a Data Provider that the authorized downstream uses of Data will be limited to computational use.*

4. No Warranty, Limitation of Liability

4.1. Data Provider does not represent or warrant that it has any rights whatsoever in the Data.

Comment: *We have chosen a broad disclaimer of representations and warranties, which may not be appropriate in commercial contexts. Because the Data Provider makes no claims that they have rights in the data, the Downstream Recipient must ensure that its use of the Data conforms to applicable laws or regulations. A Data Provider should not use the C-UDA for Data that it knows should not be distributed, or data that contains sensitive or private information.*

4.2. THE DATA IS PROVIDED ON AN “AS IS” BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, EITHER EXPRESS OR IMPLIED INCLUDING, WITHOUT LIMITATION, ANY WARRANTIES OR CONDITIONS OF TITLE, NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

4.3. NEITHER DATA PROVIDER NOR ANY UPSTREAM DATA PROVIDER SHALL HAVE ANY LIABILITY FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING WITHOUT LIMITATION LOST PROFITS), HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE DATA OR OUTPUTS, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGES.

Comment: *These disclaimers are necessary to limit Data Provider’s liability. They disclaim both express and implied warranties or representations, and the disclaimer applies to all Upstream Data Providers. These limitations of liability are common in the open data context to encourage possessors of data to share the data without requiring them to accept liability to a user for downstream uses of the data, but may not be appropriate in commercial contexts.*

5. Definitions

5.1. “Computational Use” means activities necessary to enable the use of Data (alone or along with other material) for analysis by a computer.

5.2. “Data” means the material you receive under the C-UDA in modified or unmodified form, but not including Output.

Comment: *The term “Data” encompasses both the initial data made available to “you” as well as any later modifications made to that data by a Data Provider or a Downstream Recipient that redistributes the data. This also means that Downstream Recipients remain free to modify the data. Data specifically excludes Outputs.*

5.3. “Data Provider” means the source from which you receive the Data and with whom you enter into the C-UDA.

5.4. “Downstream Recipient” means any person or persons who receives the Data directly or indirectly from you in accordance with the C-UDA.

5.5. “Output” means the outcomes or results that you obtain from your use of Data that do not include more than a de minimis portion of the Data on which the use is based. Output may include de minimis portions of the Data necessary to report on or explain use that has been conducted with the Data, such as figures in scientific papers, but do not include more. Artificial intelligence models trained on Data (and which do not include more than a de minimis portion of Data) are Output.

Comment: *The C-UDA defines “Output” to clarify that any AI model produced from the use of the Data (e.g., as a training set) should not typically be considered to be subject to the C-UDA’s restrictions. The C-UDA considers that Outputs are not a “derivative” or a “modification” of the Data if they don’t contain more than a de minimis portion of the Data. This is also intended to clarify that research papers and accompanying figures that may include only a de minimis part of the Data are not subject to any restriction in the C-UDA.*

5.6. “Upstream Data Providers” means the source or sources from which the Data Provider directly or indirectly received, under the terms of the C-UDA, material that is included in the Data.