

Data Use Agreement for Open AI Model Development

Annotated-Template

This contract template provides a set of terms suitable for governing the sharing of data by one entity with another for the purpose of allowing the second entity to use the data to ‘train’ an artificial intelligence model. Although each such scenario may involve its own set of specific considerations, there are certain core terms that lend themselves to a degree of standardization. This template aspires to present some canonical terms for community discussion that might be further refined and modified in particular circumstances.

* * *

DATA USE AGREEMENT FOR OPEN AI MODEL DEVELOPMENT

This AI Data Sharing Agreement (“Agreement”) is entered into between [●] (“Data Provider”) and [●] (“Data User”) as of [●] (“Effective Date”). Data User and Data Provider may also be referred to individually as “a party” or collectively as “the parties”.

1. Defined Terms

- a. “AI Model” means the machine learning algorithm described in Attachment A, including associated parameters and associated weights, if present.

***Comment:** This definition contemplates that the AI Model may be an entirely untrained machine learning algorithm, without any associated parameters or weights, or that it may be a partially trained algorithm. Attachment A should be used to identify and describe the particular AI Model that will be Trained by Data User.*

- b. [OPTIONAL “NDA” means a non-disclosure agreement governing the exchange of confidential information between the parties.]

- c. “Open Source License” means a license that meets all of the requirements of the “The Open Source Definition” as published by the Open Source Initiative at <https://opensource.org/osd>.

***Comment:** The definition published by the Open Source Initiative is familiar to, and generally accepted by, the open source community. Many of the most common open source licenses in use today, such as the MIT license, would satisfy the definition.*

- d. “Personal Data” means any information relating to an identified or identifiable natural person and any other information that constitutes personal data or personal information under any applicable law. An identifiable natural person is one who can be identified, directly or indirectly, in particular by referencing an identifier such as a name, an identification number, location data, an online identifier, or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.

Comment: This definition of Personal Data leverages the definition set out in the GDPR, but also extends to any other information deemed “personal data” or “personal information” by applicable law. So, for example, this definition would cover protected health information (PHI) if the data is regulated by HIPAA, because PHI is defined in 45 CFR 160.103 as “individually identifiable health information.”

- e. “Train” means to provide the AI Model with Training Data for the purpose of enhancing the predictive capabilities of the AI Model.
- f. “Trained Model” means the AI Model as modified following Training, including associated weights.
- g. “Training Data” means the dataset described in Attachment A to be provided by the Data Provider to the Data User for the purpose of Training the AI Model.

2. Provision of Data

- a. Data Provider agrees to make the Training Data (and any updates if applicable) available to the Data User as described in Attachment A.

Comment: Attachment A should be used to specify any requirements around delivery of the data, for example: when the data will be provided to the Data User, the technical mechanism to be used for that delivery, and/or any formatting requirements. If the Data Provider will be refreshing or adding to the Training Data over time, Attachment A can also make clear the nature of those updates and how frequently they will be provided to the Data User.

- b. All Personal Data (if any) included in the Training Data is as described in Attachment A.

Comment: If Personal Data is included in the Training Data, that should be made clear to Data User in Attachment A. The Parties can then use Attachment B, as referenced in Section 7c below, to set out any additional terms that may be required, e.g., to satisfy GDPR requirements.

3. Use of Data

Comment: Generally, this section includes clauses that protect the Data Provider’s interests in the Training Data.

- a. Data User agrees to use the Training Data solely for the purpose of Training the AI Model.
- b. Data User may retain the Training Data for the duration of this Agreement. Data User will delete all Training Data from its systems and records on termination of this Agreement (or based on the retention period set forth in Attachment A, if specified), except as required by applicable law.

Comment: The term of the Agreement is by default 1 year. To the extent the Data User may need to retain the data for a longer period of time, the parties can decide to extend the term of the Agreement or to add a separate (and perhaps more limited) retention right in Attachment A.

- c. Data User agrees to make Trained Model publicly available under an Open Source License that includes a general disclaimer of liability in favor of the Data Provider.

Comment: This clause requires dissemination of the Trained Model under an Open Source License in order to encourage data sharing and to facilitate the use and development of the AI model. In particular, it provides an incentive for Data Providers to share the Training Data since it would be for the benefit of the wider community. Of course, the parties to an agreement may decide to limit availability of the Trained Model in some fashion, where an Open Source License would no longer be appropriate. Note that this form does not require the Data User to maintain or update the Trained Model over time, including in the event Data Provider makes additional data available to Data User, because of the potential administrative burden that would impose, but the parties can add language to this agreement to impose such a requirement.

4. Rights Related to Trained Model

Comment: Generally, this section includes clauses that protect the Data User's ability to use the Trained Model.

- a. If there are any rights Data Provider holds in the Trained Model by virtue of Data User's Training, Data Provider irrevocably grants Data User a sublicensable license to all such rights.

Comment: This license has been included to address the possibility that a particular jurisdiction may (now or in the future) recognize the Data Provider as having some proprietary interest in the Trained Model because its data was used in the Training. It has been intentionally drafted to avoid any acknowledgment by either party that such rights exist.

- b. Data Provider has no rights or interest in any AI Models or other output developed using the Trained Model because of this Agreement. No implied rights are granted under this Agreement.

Comment: This clause provides assurance to Data User and any downstream users of the Trained Model that Data Provider has no claim to (or right to be compensated for) any intellectual property or other developments created using the Trained Model.

- c. This Agreement does not impose any restrictions with respect to the use of the Trained Model.

5. Representation and Warranties; Disclaimer

- a. Data Provider and Data User each represent and warrant that it will perform its activities in this Agreement in compliance with applicable laws, including data protection and privacy laws.

Comment: This clause gives each party assurance that the other party will be responsible for its compliance with the laws relevant to its obligations under the Agreement. For agreements subject to GDPR, the Agreement may include a provision stating that each party is an independent data controller to further clarify that each party is independently responsible for compliance.

- b. Data Provider represents and warrants that it is not aware of any contractual or other restrictions on the Training Data that would limit Data User's Training of the AI Model or use and distribution of the Trained Model as contemplated in this Agreement. Data Provider makes no representations or warranties in this Agreement with respect to Data User's rights to use and distribute the underlying AI Model.

Comment: This clause is designed to give assurance to the Data User that the Data Provider is not aware of any restrictions outside of this Agreement on the intended use of the Training Data. However, because the Data Provider is not necessarily the source of the AI Model, this clause also clarifies that the Data Provider's representations and warranties do not extend to the AI Model itself. Instead, the Data User is responsible for ensuring it has the rights it needs with respect to the AI Model, which is the subject of the next representation and warranty.

- c. Data User represents and warrants that it has sufficient rights with respect to the AI Model to Train the AI Model and distribute the Trained Model as required by this Agreement.
- d. **DATA PROVIDER MAKES NO REPRESENTATIONS OR WARRANTIES AS TO THE ACCURACY OR COMPLETENESS OF THE TRAINING DATA AND SPECIFICALLY DISCLAIMS ANY WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE WITH RESPECT TO THE TRAINING DATA. EXCEPT AS SET FORTH IN THIS SECTION 5 OR IN ATTACHMENT A, THE TRAINING DATA IS PROVIDED TO DATA USER "AS-IS" AND WITH ALL FAULTS AND DEFECTS.**

Comment: Any representations pertaining to the quality or other attributes of the data should be set out in Attachment A. The default is that the Data User uses the Training Data at its own risk.

6. Confidentiality of Training Data

- a. Data User agrees to take reasonable steps to protect the confidentiality of the Training Data while in Data User's possession or control; except that Data User may freely use or disclose any portion of the dataset described in Attachment A that: (i) was lawfully in Data User's possession prior to the time of disclosure by Data Provider; (ii) becomes publicly available without a breach of this Agreement by Data User; (iii) is received by Data User lawfully from another source without any corresponding obligation of confidentiality; or (iv) is independently developed by or for Data User.
- b. Data User may disclose the Training Data if and as required by law; but only after it notifies the Data Provider (if legally permissible) to enable the Data Provider to seek a protective order or other appropriate remedy.
- c. Data User may not disclose the Training Data to any third party, except to its employees, contractors and consultants ("Representatives") and then only on a need-to-know basis under nondisclosure obligations at least as protective as this Agreement. Data User will be responsible and liable for the use and disclosure of the Training Data by its Representatives, which use and disclosure is subject to the same limitations and requirements that apply to Data User.
- d. [OPTIONAL If there is a conflict between the terms of the NDA and the terms of this Agreement with respect to Training Data, the terms of this Agreement shall govern.]

Comment: In some instances, the parties may have entered into an NDA governing the exchange of confidential information between them. Because the Training Data is the confidential information of the Data Provider, it may be helpful to make clear in those situations that the NDA does not preclude use of the Training Data as contemplated in this Agreement.

7. Data Protection and Privacy

- a. While the Training Data is in the possession or control of Data User, Data User agrees to implement and maintain reasonable physical, administrative, and technical safeguards to protect the Training Data from inadvertent or unauthorized access, disclosure, use, or modification, taking into account the sensitivity of such Training Data.
- b. All use and storage of the Training Data by Data User will be consistent in all material respects with the data handling guidelines or frameworks set forth in Attachment A, if any.

Comment: *To the extent Data Provider requires compliance with any particular security standards, such as an ISO or NIST standard, these can be added to Attachment A.*

- c. Each party will cooperate with the other to ensure the provision, use and storage of the Training Data is in compliance with applicable laws, including any applicable data protection or privacy laws, as further described in Attachment B.

Comment: *Attachment B may be used to set out, e.g., any applicable GDPR terms. As another example, if HIPAA applies, Attachment B may take the form of a Business Associate Agreement, or may specify that the Data Provider is required to de-identify the data prior to providing it to Data User.*

- d. Data User will promptly notify Data Provider in the event of any unauthorized access, disclosure, use or modification of the Training Data and will reasonably cooperate with Data Provider to remediate and resolve such security breach to the reasonable satisfaction of Data Provider.
- e. Data User will not attempt to identify any natural person from any anonymized or de-identified Personal Data included in the Training Data.

Comment: *If desired and agreed by the parties, an audit provision could be added to the Agreement to allow Data Provider to confirm compliance with these requirements.*

8. Term and Termination

- a. This Agreement is effective as of the Effective Date and will continue until the first anniversary of the Effective Date, at which time this Agreement will automatically terminate unless extended by mutual written agreement of the parties.

Comment: *The Agreement does not automatically renew because (1) the Data Provider is likely to want a finite end to Data User's data retention rights, (2) the Data User is likely to want a finite end to its obligations under this Agreement, especially in view of the fact that the data will become stale at some point, and (3) changes in data privacy laws (or other laws) may necessitate adjustment to the contractual terms over time.*

- b. Either party may terminate this Agreement if the other party has materially breached the Agreement and has failed to cure such breach within thirty (30) days of written notification of such breach by the other party.
- c. Either party may terminate this Agreement for any reason on ninety (90) days' prior written notice to the other party.

***Comment:** Ninety days was selected to provide Data User with ample time to “wind down” use of the Training Data. This timeframe can be tailored for the particular scenario at issue.*

- d. The following Sections of this Agreement will survive termination of this Agreement: Sections 1, 3b (for the duration of the retention period, if any), 3c (for one (1) year following termination or expiration of this Agreement and thereafter until such date as Data User ceases use of the Trained Model), 4, 6, 7 (for any period during which Data User has possession or control of the Training Data), 8d and 9.

9. General

- a. **Entire Agreement; Amendments.** This Agreement is the entire agreement and understanding between the parties with respect to the subject matter described in this Agreement and supersedes all prior agreements, understandings, promises and representations with respect thereto. Any amendment to the Agreement must be in writing and is executed by authorized representatives of both parties.
- b. **Counterparts; Electronic Signatures.** This Agreement may be executed in any number of counterparts, which, when taken together, will constitute one original. This Agreement may be executed by PDF format via email or other electronically transmitted signatures and such signatures will be deemed to bind each party to this Agreement as if they were original signatures.
- c. **No Third-Party Beneficiaries.** No person or entity who is not a party to this Agreement will have the right to enforce any provision of this Agreement[, except that third party users of the Trained Model are third-party beneficiaries of Section 4(b)].

***Comment:** As a default, this form does not permit any third party to enforce the terms of this Agreement, since many of the terms relate to specific obligations as between the Data Provider and Data User (e.g., Data Provider’s obligation to provide the data, or any commitments given by the Data Provider as to quality or provenance, and Data User’s obligations with respect to retention and protection of the data). However, end users of the Trained Model should be able to invoke Section 4b in the event a Data Provider claims a proprietary interest in downstream works or other creations and so that has been added as an exception.*

- d. **Relationship of the Parties.** The parties are independent contractors and the relationship between the two parties under this Agreement will not constitute a partnership or agency. Neither party will have the authority to take any action that will be binding on the other party.
- e. **Assignment.** Neither party may assign this Agreement, in whole or in part, to any third party without the prior written consent of the other party.
- f. [OPTIONAL: **Limitations of Liability.** Except in the event of Data User’s material breach of Section 6 or unauthorized use of the Training Data: (i) in no event will either party be liable for indirect, incidental, special, punitive, or consequential damages, including loss of use, loss of profits, or interruption of business, however caused or on any theory of liability in relation to this Agreement, and (ii) to the extent permitted by applicable law, each party’s total liability for all claims relating to this Agreement will be limited to [●].]

Comment: *This optional clause should be included when the parties wish to exclude the recovery of indirect damages and/or to cap liability at a specific dollar amount. This clause has not been made mandatory as the nature and sensitivity of the Training Data may drive parties to adopt different approaches. There may also be instances in which the parties may wish to include the provision but to make it one-way in favor of Data Provider, e.g., where the Data Provider is providing the data strictly “as is” but could be exposed to losses if the data is misused by Data User.*

Reviewed and Agreed

On behalf of [DATA PROVIDER]

By:

Name:

Title:

On behalf of [DATA USER]

By:

Name:

Title:

ATTACHMENT A
Project Details

Part I - Description of AI Model; License

Insert description of AI Model (including, if publicly accessible, the relevant URL), the source of the AI Model (if known), and identify the license (if applicable) under which it is made available to Data User. Specify any copyright notices or attribution requirements.

Part II - Description of Training Data

Insert description of Training Data and, if desirable, any limitations/disclaimers about the data (e.g., if it based on a subset of a particular population, collected during a specified time period, known to be incomplete, etc.).

[Optional: Insert information about the provenance and lineage of the Training Data, as well as any legal, contractual, or other limitations on Data Provider's rights to transfer, process or otherwise use, or permit others to transfer, process or otherwise use, the Training Data.]

[Optional: Insert any certifications as to the contents, quality, or other characteristics of the Training Data that should be a carve-out to the general disclaimer in Section 5d.]

[Optional: Insert any description of the methods/technologies used by the Data Provider to secure the data as well as any expectations about the Data User's obligations to provide the same or equivalent protection.

Part III - AI Project Specification

Insert description of delivery mechanism for the Training Data as well as any formatting requirements and (if applicable) the frequency of updates.

[Optional: Insert data retention period (i.e., the duration of time Data User is permitted to retain the Training Data following termination/expiration of the Agreement).]

[Optional: Insert any applicable data handling guidelines or frameworks that should apply to Data User's storage and use of the Training Data (e.g., ISO/NIST standards).]

Part IV - Location of Trained Model

Insert URL where Data User will make the Trained Model publicly available for download, and include any other relevant instructions for accessing the model.

ATTACHMENT B

Data Privacy

If relevant, insert applicable GDPR, HIPAA or other applicable privacy terms.

Comment: *Generally, this Attachment B allows the Data Provider to specify data handling and treatment practices to preserve the privacy of the data subjects from which the data was collected. For example, if the Data Provider has aggregated or pseudonymized the data in an effort to address privacy concerns, the Data User may be required to adhere to certain data handling practices to ensure that those steps are not circumvented either intentionally or unintentionally. Also, the parties may want to limit human inspection of some or all of the data. In other circumstances, 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. Each of these approaches may require specific data handling practices to be effective.*