



Dataset Name (Acronym)	Write a short summary describing your dataset (limit 200 words). Include information about the content and topic of the data, sources and motivations for the dataset, benefits and the problems or use cases it is suitable for.
DATASET LINK	DATA CARD AUTHOR(S)
<i>Provide a link to the dataset:</i>	<i>Select one role per Data Card Author:</i> <i>(Usage Note: Select the most appropriate choice to describe the author's role in creating the Data Card.)</i>
Dataset Link	Name, Team: (Owner / Contributor / Manager) Name, Team: (Owner / Contributor / Manager) Name, Team: (Owner / Contributor / Manager)

Authorship		
Publishers		
PUBLISHING ORGANIZATION(S)	INDUSTRY TYPE(S)	CONTACT DETAIL(S)
<i>Provide the names of the institution or organization responsible for publishing the dataset:</i>	<i>Select all applicable industry types to which the publishing organizations belong:</i>	<i>Provide publisher contact details:</i>
Organization Name	Corporate - Tech Corporate - Non-Tech (please specify) Academic - Tech Academic - Non-Tech (please specify) Not-for-profit - Tech	Publishing POC: <Provide the name for a POC for this dataset's publishers.> Affiliation: <Provide the POC's institutional affiliation.> Contact: <Provide the POC's contact details.> Mailing List: <Provide a mailing list if available.>

	Not-for-profit - Non-Tech (please specify) Individual (please specify) Others (please specify)	Website: <Provide a website for the dataset if available.>
Dataset Owners		
TEAM(S)	CONTACT DETAIL(S)	AUTHOR(S)
<i>Provide the names of the groups or team(s) that own the dataset:</i>	<i>Provide pathways to contact dataset owners:</i>	<i>Provide the details of all authors associated with the dataset:</i> <i>(Usage Note: Provide the affiliation and year if different from publishing institutions or multiple affiliations.)</i>
Name of Group or Team	Dataset Owner(s): <Provide the names of the dataset owners> Affiliation: <Provide the affiliation of the dataset owners> Contact: <Provide the email of the dataset owner> Group Email: <Provide a link to the Mailing-list@server.com for the dataset owner team> Website: <Provide a link to the website for the dataset owner team>	Name, Title, Affiliation, YYYY Name, Title, Affiliation, YYYY Name, Title, Affiliation, YYYY Name, Title, Affiliation, YYYY
Funding Sources		
INSTITUTION(S)	FUNDING OR GRANT SUMMARY(IES)	
<i>Provide the names of the funding institution(s):</i>	<i>Provide a short summary of programs or projects that may have funded the creation, collection, or curation of the dataset.</i> <i>Use additional notes to capture any other relevant information or considerations.</i>	
Name of Institution Name of Institution Name of Institution	<i>For example, Institution 1 and institution 2 jointly funded this dataset as a part of the XYZ data program, funded by XYZ grant awarded by institution 3 for the years YYYY-YYYY.</i> <Summarize here. Link to documents if available.> Additional Notes: <Add here>	

Dataset Overview

DATA SUBJECT(S)

Select **all applicable** subjects contained the dataset:

Sensitive Data about people

Non-Sensitive Data about people

Data about natural phenomena

Data about places and objects

Synthetically generated data

Data about systems or products and their behaviors

Unknown

Others (Please specify)

Size of Dataset

Number of Instances

Number of Fields

Labeled Classes

Number of Labels

Average Labels Per Instance

Algorithmic Labels

Human Labels

Other Characteristics

123456 MB

123456

123456

123456

123456789

123456

123456789

123456789

123456

Above: <Provide a caption for the above table or visualization.>

Additional Notes: <Add here>

DATASET SNAPSHOT

Provide a snapshot of the dataset:

(Use the additional notes to include relevant information, considerations, and links to table(s) with more detailed breakdowns.)

CONTENT DESCRIPTION

Provide a short description of the content in a datapoint.

<Summarize here. Include links if available>

Additional Notes: <Add here>

DESCRIPTIVE STATISTICS

Provide basic descriptive statistics for each field.

Use additional notes to capture any other relevant information or considerations.

Usage Note: Some statistics will be relevant for numeric data, but not for strings.

Statistic	Field Name	Field Name	Field Name	Field Name	Field Name	Field Name
count						
mean						
std						
min						
25%						
50%						

75%						
max						
mode						

Above: <Provide a caption for the above table or visualization.>

Additional Notes: <Add here>

Sensitivity of Data

SENSITIVITY TYPE(S)	FIELD(S) WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING														
<p><i>Select all applicable data types present in the dataset:</i></p>	<p><i>List fields in the dataset that contain S/PII, and specify if their collection was intentional or unintentional.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>	<p><i>Summarize the measures or steps to handle sensitive data in this dataset.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>														
<div>User Content</div> <div>User Metadata</div> <div>User Activity Data</div> <div>Identifiable Data</div> <div>S/PII</div> <div>Business Data</div> <div>Employee Data</div> <div>Pseudonymous Data</div> <div>Anonymous Data</div> <div>Health Data</div> <div>Children's Data</div> <div>None</div> <div>Others (Please specify)</div>	<div>Intentionally Collected Sensitive Data</div> <div>(S/PII were collected as a part of the dataset creation process.)</div> <table> <tr> <th>Field Name</th> <th>Description</th> </tr> <tr> <td><Field Name></td> <td><Type of S/PII></td> </tr> <tr> <td><Field Name></td> <td><Type of S/PII></td> </tr> <tr> <td><Field Name></td> <td><Type of S/PII></td> </tr> <tr> <td>...</td> <td>...</td> </tr> </table> <div>Unintentionally Collected Sensitive Data</div> <div>(S/PII were not explicitly collected as a part of the dataset creation process but can be inferred using additional methods.)</div> <table> <tr> <th>Field Name</th> <th>Description</th> </tr> <tr> <td><Field Name></td> <td><Type of S/PII></td> </tr> </table>	Field Name	Description	<Field Name>	<Type of S/PII>	<Field Name>	<Type of S/PII>	<Field Name>	<Type of S/PII>	Field Name	Description	<Field Name>	<Type of S/PII>	<div><Summarize here. Include links and metrics where applicable.></div> <div><Method>: [description]</div> <div><Method>: [description]</div> <div><Method>: [description]</div> <div>Additional Notes: <Add here></div>
Field Name	Description															
<Field Name>	<Type of S/PII>															
<Field Name>	<Type of S/PII>															
<Field Name>	<Type of S/PII>															
...	...															
Field Name	Description															
<Field Name>	<Type of S/PII>															

	<table> <tr> <td><Field Name></td><td><Type of S/PII></td></tr> <tr> <td><Field Name></td><td><Type of S/PII></td></tr> <tr> <td>...</td><td>...</td></tr> </table> <p>Additional Notes: <Add here></p>	<Field Name>	<Type of S/PII>	<Field Name>	<Type of S/PII>	
<Field Name>	<Type of S/PII>							
<Field Name>	<Type of S/PII>							
...	...							
RISK TYPE(S)	SUPPLEMENTAL LINK(S)	RISK(S) AND MITIGATION(S)						
<p>Select all applicable risk types presenting from the dataset:</p>	<p>Provide link(s) for documentation pertaining to sensitive data in the dataset:</p>	<p>Summarize the steps taken to identify and mitigate risks from PII or sensitive information.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>						
<p>Direct Risk</p> <p>Indirect Risk</p> <p>Residual Risk</p> <p>No Known Risks</p> <p>Others (Please Specify)</p>	<p><Link Name or Document Type>: [Link]</p> <p><Link Name or Document Type>: [Link]</p> <p><Link Name or Document Type>: [Link]</p>	<p><Summarize here. Include links and metrics where applicable.></p> <p><Risk type>: [Description + Mitigations]</p> <p><Risk type>: [Description + Mitigations]</p> <p><Risk type>: [Description + Mitigations]</p> <p>Additional Notes: <Add here></p>						
Dataset Version and Maintenance								
MAINTENANCE STATUS	VERSION DETAILS	MAINTENANCE PLAN						
<p>Select one:</p>	<p>Provide details about this version of the dataset:</p>	<p>Summarize the maintenance plan for the dataset:</p> <p>Use additional notes to capture any other relevant information or considerations.</p>						
<p>Regularly Updated (New versions of the dataset have been or will continue to be made available.)</p> <p>Actively Maintained (No new versions will be made available, but this dataset will be actively maintained, including but not limited to updates to the data.)</p>	<p>Current Version: 1.0</p> <p>Last Updated: MM/YYYY</p> <p>Release Date: MM/YYYY</p>	<p><Summarize here. Include links and metrics where applicable.></p> <p>Versioning: <Summarize here. Include information about criteria for versioning the dataset.></p> <p>Updates: <Summarize here. Include information about criteria for refreshing or updating the dataset.></p> <p>Errors: <Summarize here. Include information about how errors are triaged or handled.></p>						

<p>Limited Maintenance (The data will not be updated, but any technical issues will be addressed.)</p> <p>Deprecated (This dataset is obsolete or is no longer being maintained.)</p>		<p>Feedback: <Summarize here. Include information for providing feedback.></p> <p>Additional Notes: <Add here></p>
	NEXT PLANNED UPDATE(S)	EXPECTED CHANGE(S)
<i>Fill this row if this dataset is regularly updated, actively maintained or another version is planned.</i>	<i>Provide details about the next planned update:</i>	<i>Summarize the updates to the dataset and/or data that are expected on the next update.</i> <i>Use additional notes to capture any other relevant information or considerations.</i>
	<p>Version affected: 1.0</p> <p>Next data update: MM/YYYY</p> <p>Next Version: 1.1</p> <p>Next Version update: MM/YYYY</p>	<p>Updates to Data: <Summarize here. Include links, charts, and visualizations as appropriate.></p> <p>Updates to Dataset: <Summarize here. Include links, charts, and visualizations as appropriate.></p> <p>Additional Notes: <Add here></p>

Example of Data Points

PRIMARY DATA MODALITY

Select **one**:

Image Data

Text Data

Tabular Data

Audio Data

Video Data

Time Series

Graph Data

Geospatial Data

Multimodal (Please Specify)

Unknown

Others (Please specify)

SAMPLING OF DATA POINTS

[Demo Link]

[Typical Data Point Link]

[Outlier Data Point Link]

[Other Datapoint Link]

[Other Datapoint Link]

DATA FIELDS

List the fields in data points and their descriptions.

(Usage Note: Describe each field in a datapoint. Optionally use this to show the example.)

Field Name	Field Value	Description
<Field Name>	<Field Value>	<Description >
<Field Name>	<Field Value>	<Description >
<Field Name>	<Field Value>	<Description >
...

Above: <Provide a caption for the above table or visualization if used.>

Additional Notes: <Add here>

TYPICAL DATA POINT

Provide an example of a typical data point and describe what makes it typical.

Use additional notes to capture any other relevant information or considerations.

ATYPICAL DATA POINT

Provide an example of an outlier data point and describe what makes it atypical.

Use additional notes to capture any other relevant information or considerations.

<Summarize here. Include any criteria for typicality of datapoint>

```

{
  'q_id': '8houtx',
  'title': 'Why does water heated to room temperature feel colder than the air around it?',
  'selftext': '',
  'document': '',
  'subreddit': 'explainlikeimfive',
}

```

```
'subreddit': 'explainlikeimfive',
'answers': {'a_id': ['dylcnfk', 'dylcj49'],
'text': ["Water transfers heat more
efficiently than air. When something feels
cold it's because heat is being transferred
from your skin to whatever you're touching.
... Get out of the water and have a breeze
blow on you while you're wet, all of the water
starts evaporating, pulling even more heat
from you."],
'score': [5, 2]},
'title_urls': {'url': []},
'selftext_urls': {'url': []},
'answers_urls': {'url': []}} ``
```

Additional Notes: <Add here>

```
'answers': {'a_id': ['dylcnfk',
'dylcj49'],
'text': ["Water transfers heat more
efficiently than air. When something
feels cold it's because heat is being
transferred from your skin to whatever
you're touching....Get out of the water
and have a breeze blow on you while
you're wet, all of the water starts
evaporating, pulling even more heat from
you."],
'score': [5, 2]},
'title_urls': {'url': []},
'selftext_urls': {'url': []},
'answers_urls': {'url': []}} ``
```

Additional Notes: <Add here>

Motivations & Intentions		
Motivations		
PURPOSE(S)	DOMAIN(S) OF APPLICATION	MOTIVATING FACTOR(S)
<p>Select one:</p>	<p><i>Provide a list of key domains of application that the dataset has been designed for:</i></p> <p><i>(Usage Note: Use comma-separated keywords.)</i></p>	<p><i>List the primary motivations for creating or curating this dataset:</i></p> <p><i>(Usage Note: use this to describe the problem space and corresponding motivations for the dataset.)</i></p>
<p>Monitoring</p> <p>Research</p> <p>Production</p> <p>Others (Please Specify)</p>	<p><i>For example: `Machine Learning`, `Computer Vision`, `Object Detection`.</i></p> <p><i>`keyword`, `keyword`, `keyword`</i></p>	<p><i>For example:</i></p> <ul style="list-style-type: none"> <i>- Bringing demographic diversity to imagery training data for object-detection models.</i> <i>- Encouraging academics to take on second-order challenges of cultural representation in object detection.</i> <p><Summarize motivation here. Include links where relevant.></p>
Intended Use		
DATASET USE(S)	SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
<p>Select one:</p>	<p><i>Summarize known suitable and intended use cases of this dataset.</i></p> <p><i>Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.</i></p>	<p><i>Summarize known unsuitable and unintended use cases of this dataset.</i></p> <p><i>Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.</i></p>
<p>Safe for production use</p> <p>Safe for research use</p> <p>Conditional use-some unsafe applications</p> <p>Only approved use</p> <p>Others (Please specify)</p>	<p>[Suitable Use Case] : <Summarize here. Include links where necessary.></p> <p>[Suitable Use Case]: <Summarize here. Include links where necessary.></p> <p>[Suitable Use Case]: <Summarize here. Include links where necessary.></p> <p>Additional Notes: <Add here></p>	<p>[Unsuitable Use Case] : <Summarize here. Include links where necessary.></p> <p>[Unsuitable Use Case]: <Summarize here. Include links where necessary.></p> <p>[Unsuitable Use Case]: <Summarize here. Include links where necessary.></p> <p>Additional Notes: <Add here></p>
	RESEARCH AND PROBLEM SPACE(S)	CITATION GUIDELINES

	<p><i>Provide a description of the specific problem space that this dataset intends to address.</i></p>	<p><i>Provide guidelines and steps for citing this dataset in research and/or production work.</i></p> <p><i>Use additional notes to capture any specific patterns that readers should look out for, or other relevant information or considerations.</i></p>
	<p><Summarize here. Include any specific research questions.></p>	<p>Guidelines & Steps: <Summarize here. Include links where necessary.></p> <p>BiBTeX:</p> <pre> ``` @article{kuznetsova2020open, title={The open images dataset v4}, author={Kuznetsova, Alina and Rom, Hassan and Alldrin, and others}, journal={International Journal of Computer Vision}, volume={128}, number={7}, pages={1956--1981}, year={2020}, publisher={Springer} } ``` </pre> <p>Additional Notes: <Add here></p>

Access, Retention, & Wipeout

Access

ACCESS TYPE

DOCUMENTATION LINK(S)

PREREQUISITE(S)

Select **one**:

Provide links that describe documentation to access this dataset:

Please describe any required training or prerequisites to access this dataset.

Internal - Unrestricted

[Dataset Website URL]

Internal - Restricted

[Github URL]

External - Open Access

Others (Please specify)

For example,
This dataset requires membership in [specific] database groups:

- Complete the [Mandatory Training]
- Read [Data Usage Policy]
- Initiate a Data Requesting by filing [a bug]

POLICY LINK(S)

ACCESS CONTROL LIST(S)

List and summarize any access control lists associated with this dataset. Include links where necessary.

Use additional notes to capture any other information relevant to accessing the dataset.

Provide a link to the access policy:

- Direct download URL
- Other repository URL

Code to download data

...

[Access Control List]: <Write summary and notes here.>

[Access Control List]: <Write summary and notes here.>

[Access Control List]: <Write summary and notes here.>

Additional Notes: <Add here>

Retention

DURATION

POLICY SUMMARY

Specify the duration for which this dataset can be retained:

Summarize the retention policy for this dataset.

<Specify duration in days, months, or years.>

Retention Plan ID: <Write here>

Summary: <write summary and notes here>

PROCESS GUIDE

EXCEPTION(S) AND EXEMPTION(S)

	<p><i>Summarize any requirements and related steps to retain the dataset.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>	<p><i>Summarize any exceptions and related steps to retain the dataset. Include links where necessary.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>
	<p><i>For example, This dataset complies with [standard policy guidelines]</i></p> <p>Additional Notes: <Add here></p>	<p>Exemption Code: `ANONYMOUS_DATA` / `EMPLOYEE_DATA` / `PUBLIC_DATA` / `INTERNAL_BUSINESS_DATA` / `SIMULATED_TEST_DATA`</p> <p>Summary: <Write summary and notes here.></p> <p>Additional Notes: <Add here></p>
Wipeout and Deletion		
	DURATION	DELETION EVENT SUMMARY
	<p><i>Specify the duration after which this dataset should be deleted or wiped out:</i></p>	<p><i>Summarize the sequence of events and allowable processing for data deletion.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>
	<p><Specify duration in days, months, or years.></p>	<p>Sequence of deletion and processing events:</p> <p><Summarize first event here.></p> <p><Summarize second event here.></p> <p><Summarize third event here.></p> <p>Additional Notes: <Add here></p>
	ACCEPTABLE MEANS OF DELETION	POST-DELETION OBLIGATIONS
	<p><i>List the acceptable means of deletion:</i></p>	<p><i>Summarize the sequence of obligations after a deletion event.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>
	<p><Write acceptable means of deletion.></p> <p><Write acceptable means of deletion.></p>	<p>Sequence of post-deletion obligations:</p> <p><Summarize first obligation here.></p> <p><Summarize second obligation here.></p> <p><Summarize third obligation here.></p>

	<Write acceptable means of deletion.>	Additional Notes: <Add here>
	OPERATIONAL REQUIREMENT(S)	EXCEPTIONS AND EXEMPTIONS
	<i>List any wipeout integration operational requirements:</i>	<i>Summarize any exceptions and related steps to a deletion event.</i> <i>Use additional notes to capture any other relevant information or considerations.</i>
	Wipeout Integration Operational Requirements: <Write first requirement here.> <Write second requirement here.> <Write third requirement here.>	Policy Exception bug: [bug] Summary: <Write summary and notes here.> Additional Notes: <Add here>

Provenance

Collection

METHOD(S) USED	METHODOLOGY DETAIL(S)	SOURCE DESCRIPTION(S)
<p>Select all applicable methods used to collect data:</p>	<p>Provide a description of each collection method used.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for collection method type.)</p>	<p>Provide a description of each upstream source of data.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
<p>API</p> <p>Artificially Generated</p> <p>Crowdsourced - Paid</p> <p>Crowdsourced - Volunteer</p> <p>Vendor Collection Efforts</p> <p>Scraped or Crawled</p> <p>Survey, forms or polls</p> <p>Taken from other existing datasets</p> <p>Unknown</p> <p>To be determined</p> <p>Others (Please specify)</p>	<p><Collection Type></p> <p>Source: <Describe here. Include links where available.></p> <p>Platform: [Platform Name], <Describe platform here. Include links where relevant.></p> <p>Is this source considered sensitive or high-risk? [Yes / No]</p> <p>Dates of Collection: [MMM YYYY - MMM YYYY]</p> <p>Primary modality of collected data: <i>Usage Note: Select one for this collection type.</i> Image Data Text Data Tabular Data Audio Data Video Data Time Series Graph Data Geospatial Data Unknown Multimodal (Please specify) Others (Please specify)</p> <p>Update Frequency for collected data: <i>Usage Note: Select one for this collection type.</i> Yearly Quarterly Monthly Biweekly Weekly Daily Hourly Static Others (Please specify)</p> <p>Additional Links for this collection:</p> <ul style="list-style-type: none"> • [Access Policy] • [Wipeout Policy] • [Retention Policy] 	<p>[Source]: <Describe here. Include links, data examples, metrics, visualizations where relevant.></p> <p>[Source]: <Describe here. Include links, data examples, metrics, visualizations where relevant.></p> <p>[Source]: <Describe here. Include links, data examples, metrics, visualizations where relevant.></p> <p>Additional Notes: <Add here></p>

	Additional Notes: <Add here>																	
COLLECTION CADENCE	DATA INTEGRATION	DATA PROCESSING																
	<p><i>List all fields collected from different sources, and specify if they were included or excluded from the dataset.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate and complete the following for each upstream source.)</i></p>	<p><i>Summarize how data from different sources or methods aggregated, processed, or connected.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate and complete the following for each source OR collection method.)</i></p>																
<p>Select <i>all applicable</i>:</p> <p>Static (Data was collected once from single or multiple sources.)</p> <p>Streamed (Data is continuously acquired from single or multiple sources.)</p> <p>Dynamic (Data is updated regularly from single or multiple sources.)</p> <p>Others (Please specify)</p>	<p><Source> Included Fields (Data fields that were collected and are included in the dataset.)</p> <table><tr><th>Field Name</th><th>Description</th></tr><tr><td><Field Name></td><td><Describe here. Include links, data examples, metrics, visualizations where relevant.></td></tr><tr><td><Field Name></td><td><Describe here. Include links, data examples, metrics, visualizations where relevant.></td></tr><tr><td>...</td><td>...</td></tr></table> <p>Additional Notes: <Add here></p> <p>Excluded Fields (Data fields that were collected but are excluded from the dataset.)</p> <table><tr><th>Field Name</th><th>Description</th></tr><tr><td><Field Name></td><td><Describe here. Include links, data examples, metrics, visualizations where relevant.></td></tr><tr><td><Field Name></td><td><Describe here. Include links, data examples, metrics, visualizations where relevant.></td></tr><tr><td>...</td><td>...</td></tr></table>	Field Name	Description	<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>	<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>	Field Name	Description	<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>	<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>	<p><Collection Method or Source></p> <p>Description: <Describe here. Include links where relevant.></p> <p>Methods employed: <Describe here. Include links where relevant.></p> <p>Tools or libraries: <Describe here. Include links where relevant.></p> <p>Additional Notes: <Add here></p>
Field Name	Description																	
<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>																	
<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>																	
...	...																	
Field Name	Description																	
<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>																	
<Field Name>	<Describe here. Include links, data examples, metrics, visualizations where relevant.>																	
...	...																	

	Additional Notes: <Add here>	
Collection Criteria		
DATA SELECTION	DATA INCLUSION	DATA EXCLUSION
<i>Summarize the data selection criteria.</i>		<i>Summarize the data exclusion criteria.</i>
<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Summarize the data inclusion criteria.</i>	<i>Use additional notes to capture any other relevant information or considerations.</i>
<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Use additional notes to capture any other relevant information or considerations.</i>
[Collection Method or Source]: <Summarize data selection criteria here. Include links where available.> [Collection Method or Source]: <Summarize data selection criteria here. Include links where available.> [Collection Method or Source]: <Summarize data selection criteria here. Include links where available.> [Collection Method or Source]: <Summarize data selection criteria here. Include links where available.> [Collection Method or Source]: <Summarize data selection criteria here. Include links where available.> Additional Notes: <Add here>	[Collection Method or Source]: <Summarize data inclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data inclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data inclusion criteria here. Include links where available.> Additional Notes: <Add here>	[Collection Method or Source]: <Summarize data exclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data exclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data exclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data exclusion criteria here. Include links where available.> [Collection Method or Source]: <Summarize data exclusion criteria here. Include links where available.> Additional Notes: <Add here>
Relationship to Source		
USE & UTILITY(IES)	BENEFIT AND VALUE(S)	LIMITATION(S) AND TRADE-OFF(S)
<i>Describe how the resulting dataset is aligned with the purposes, motivations, or intended use of the upstream source(s).</i>		
<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Summarize the benefits of the resulting dataset to its consumers, compared to the upstream source(s).</i>	<i>What are the limitations of the resulting dataset to its consumers, compared to the upstream source(s)?</i>
<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Break down by source type.</i>
<i>(Usage Note: Duplicate and complete the following for each source type.)</i>	<i>(Usage Note: Duplicate and complete the following for each source type.)</i>	<i>(Usage Note: Duplicate and complete the following for each source type.)</i>
[Source Type]: <Summarize here. Include links where available.>	[Source Type]: <Summarize here. Include links where available.> [Source Type]: <Summarize here. Include links where available.>	[Source Type]: <Summarize here. Include links where available.>

<p>[Source Type]: <Summarize here. Include links where available.></p> <p>[Source Type]: <Summarize here. Include links where available.></p> <p>Additional Notes: <Add here></p>	<p>[Source Type]: <Summarize here. Include links where available.></p> <p>Additional Notes: <Add here></p>	<p>[Source Type]: <Summarize here. Include links where available.></p> <p>[Source Type]: <Summarize here. Include links where available.></p>
<h3>Version and Maintenance</h3>		
<p>Fill this next row if: this is not the first version of the dataset, and there is no data card available for the first version.</p>		
	FIRST VERSION	NOTE(S) AND CAVEAT(S)
	<p>Provide a basic description of the first version of this dataset.</p>	<p>Summarize the caveats or nuances of the first version of this dataset that may affect the use of the current version.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
	<p>Release date: MM/YYYY</p> <p>Link to dataset: [Dataset Name + Version]</p> <p>Status: [Select one: Actively Maintained/ Limited Maintenance / Deprecated]</p> <p>Size of Dataset: 123 MB</p> <p>Number of Instances: 123456</p>	<p><Summarize here. Include links where available.></p> <p>Additional Notes: <Add here></p>
CADENCE	LAST AND NEXT UPDATE(S)	CHANGES ON UPDATE(S)
<p>Select one.</p>	<p>Please describe the update schedule</p>	<p>Summarize the changes that occur when the dataset is refreshed.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each source type.)</p>

Yearly	Date of last update: DD/MM/YYYY Total Data points affected: 12345 Data points updated: 12345 Data points added: 12345 Data points removed: 12345 Date of next update: DD/MM/YYYY	[Source Type]: <Summarize here. Include links where available.>
Quarterly		[Source Type]: <Summarize here. Include links where available.>
Monthly		[Source Type]: <Summarize here. Include links where available.>
Biweekly		[Source Type]: <Summarize here. Include links where available.>
Weekly		Additional Notes: <Add here>
Daily		
Hourly		
Static		
Others (Please Specify)		

Human and Other Sensitive Attributes																		
SENSITIVE HUMAN ATTRIBUTE(S)	INTENTIONALITY	RATIONALE																
<p>Select all attributes that are represented (directly or indirectly) in the dataset.</p> <p>[Note: there are demonstrative examples of potentially relevant signals for a given dataset analysis]</p>	<p>List fields in the dataset that contain human attributes, and specify if their collection was intentional or unintentional.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>	<p>Describe the motivation, rationale, considerations or approaches that caused this dataset to include the indicated human attributes. Summarize why or how this might affect the use of the dataset.</p>																
<p>Gender</p> <p>Socio-economic status</p> <p>Geography</p> <p>Language</p> <p>Age</p> <p>Culture</p> <p>Experience or Seniority</p> <p>Others (Please Specify)</p>	<div><p>Intentionally Collected Attributes (Human attributes were labeled or collected as a part of the dataset creation process.)</p><table><tr><th>Field Name</th><th>Description</th></tr><tr><td><Field Name></td><td><Human Attributed Collected.></td></tr><tr><td><Field Name></td><td><Human Attributed Collected.></td></tr><tr><td>...</td><td>...</td></tr></table><p>Additional Notes: <Add here></p><p>Unintentionally Collected Attributes (Human attributes were not explicitly collected as a part of the dataset creation process but can be inferred using additional methods.)</p><table><tr><th>Field Name</th><th>Description</th></tr><tr><td><Field Name></td><td><Human Attributed Collected.></td></tr><tr><td><Field Name></td><td><Human Attributed Collected.></td></tr><tr><td>...</td><td>...</td></tr></table><p>Additional Notes: <Add here></p></div>	Field Name	Description	<Field Name>	<Human Attributed Collected.>	<Field Name>	<Human Attributed Collected.>	Field Name	Description	<Field Name>	<Human Attributed Collected.>	<Field Name>	<Human Attributed Collected.>	<p><Summarize here. Include links, tables, and media as relevant></p>
Field Name	Description																	
<Field Name>	<Human Attributed Collected.>																	
<Field Name>	<Human Attributed Collected.>																	
...	...																	
Field Name	Description																	
<Field Name>	<Human Attributed Collected.>																	
<Field Name>	<Human Attributed Collected.>																	
...	...																	
	SOURCE(S)	METHODOLOGY DETAIL(S)																
	<p>List the sources of the human attributes.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>	<p>Describe the methods used to collect human attributes in the dataset.</p> <p>Use additional notes to capture any other relevant information or considerations. (Usage Note: Duplicate and complete the following for each human attribute.)</p>																

	[Human Attribute]: Sources [Human Attribute]: Sources [Human Attribute]: Sources Additional Notes: <Add here>	[Human Attribute] Method: <Describe the collection method here. Include links where necessary.> Collection task: <Describe the task here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> Additional Notes: <Add here>
--	--	--

DISTRIBUTION(S)

Provide basic descriptive statistics for each human attribute, noting key takeaways in the caption. Use additional notes to capture any other relevant information or considerations. (Usage Note: Duplicate and complete the following for each human attribute.)

Human Attribute

[Human attribute]	Label or Class	Label or Class	Label or Class	Label or Class
Count	123456	123456	123456	123456
[Statistic]	123456	123456	123456	123456
[Statistic]	123456	123456	123456	123456
[Statistic]	123456	123456	123456	123456

Above: <Provide a caption for the above table or visualization.>
Additional Notes: <Add here>

	KNOWN CORRELATIONS	RISK(S) AND MITIGATION(S)
	<p>Describe any known correlations with the indicated sensitive attributes in this dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate for each known correlation.)</p>	<p>Summarize systemic or residual risks, performance expectations, trade-offs and caveats because of human attributes in this dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>Usage Note: Duplicate and complete the following for each human attribute.)</p>

	<p>[`field_name`,`field_name`]</p> <p>Description: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Impact on dataset use: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Additional Notes: <Add here></p>	<p>[Human Attribute]</p> <p><Summarize here. Include links and metrics where applicable.></p> <p><Risk type>: [Description + Mitigations]</p> <p><Risk type>: [Description + Mitigations]</p> <p><Risk type>: [Description + Mitigations]</p> <p>Trade-offs, caveats, & other considerations: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Additional Notes: <Add here></p>
--	---	--

Extended Use

Use with Other Data

SAFETY LEVEL	KNOWN SAFE DATASET(S) OR DATA TYPE(S)	BEST PRACTICES
<p>Select one:</p> <p>Safe to use with other data</p> <p>Conditionally safe to use with other data</p> <p>Should not be used with other data</p> <p>Unknown</p> <p>Others (Please Specify)</p>	<p>List the known datasets or data types and corresponding transformations that are safe to join or aggregate this dataset with.</p>	<p>Summarize best practices for using this dataset with other datasets or data types.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
	<p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p>	<p><Summarize here. Include visualizations, metrics, demonstrative examples, or links where necessary.></p> <p>Additional Notes: <Add here></p>
	KNOWN UNSAFE DATASET(S) OR DATA TYPE(S)	LIMITATION(S) AND RECOMMENDATION(S)
<p>Fill out this row if you selected “Conditionally safe to use with other datasets” or “Should not be used with other datasets”:</p>	<p>List the known datasets or data types and corresponding transformations that are unsafe to join or aggregate with this dataset.</p>	<p>Summarize limitations of the dataset that introduce foreseeable risks when the dataset is conjoined with other datasets.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
	<p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Dataset or Data Type: <Summarize here. Include visualizations, metrics, or links where necessary.></p>	<p><Summarize here. Include links and metrics where applicable.></p> <p><Limitation type>: [Dataset or data type, description and recommendation.]</p> <p><Limitation type>: [Dataset or data type, description and recommendation.]</p> <p><Limitation type>: [Dataset or data type, description and recommendation.]</p> <p>Additional Notes: <Add here></p>

Forking & Sampling		
SAFETY LEVEL	ACCEPTABLE SAMPLING METHOD(S)	BEST PRACTICE(S)
<p>Select one:</p> <p>Safe to form and/or sample</p> <p>Conditionally safe to fork and/or sample</p> <p>Should not be forked and/or sampled</p> <p>Unknown</p> <p>Others (Please specify)</p>	<p>Select all applicable acceptable methods to sample this dataset:</p> <p>Cluster Sampling</p> <p>Haphazard Sampling</p> <p>Multi-stage Sampling</p> <p>Random Sampling</p> <p>Retrospective Sampling</p> <p>Stratified Sampling</p> <p>Systematic Sampling</p> <p>Weighted Sampling</p> <p>Unknown</p> <p>Unsampled</p> <p>Others (Please specify)</p>	<p>Summarize the best practices for forking or sampling this dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p><Summarize here. Include links, figures and demonstrative examples where available.></p> <p>Additional Notes: <Add here></p>
	RISK(S) AND MITIGATION(S)	LIMITATION(S) AND RECOMMENDATION(S)
<p>Fill out this row if you selected “Conditionally safe to fork and/or sample” or “Should not be forked and/or sampled”.</p>	<p>Summarize known or residual risks associated with forking and sampling methods when applied to the dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>	<p>Summarize the limitations that the dataset introduces when forking or sampling the dataset and corresponding recommendations.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
	<p><Summarize here. Include links and metrics where applicable.></p> <p><Risk Type>: [Description + Mitigations]</p> <p><Risk Type>: [Description + Mitigations]</p> <p><Risk Type>: [Description + Mitigations]</p> <p>Additional Notes: <Add here></p>	<p><Summarize here. Include links and metrics where applicable.></p> <p><Limitation Type>: [Description + Recommendation]</p> <p><Limitation Type>: [Description + Recommendation]</p> <p><Limitation Type>: [Description + Recommendation]</p> <p>Additional Notes: <Add here></p>
Use in ML or AI Systems		
DATASET USE(S)	NOTABLE FEATURE(S)	USAGE GUIDELINE(S)

	<p><i>Describe any notable feature distributions or relationships between individual instances made explicit.</i></p> <p><i>Include links to servers where readers can explore the data on their own.</i></p>	<p><i>Summarize usage guidelines or policies that consumers should be aware of.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>								
Select all applicable :										
Training	Exploration Demo: [Link to server or demo.]	Usage Guidelines: <Summarize here. Include links where necessary.>								
Testing	<Notable Field Name>: <Describe here. Include links, data examples, metrics, visualizations where relevant.>	Approval Steps: <Summarize here. Include links where necessary.>								
Validation	Above: <Provide a caption for the above table or visualization.>	Reviewer: <Provide the name of a reviewer for publications referencing this dataset.>								
Development or Production Use	Additional Notes: <Add here>	Additional Notes: <Add here>								
Fine Tuning										
Others (Please Specify)										
	DISTRIBUTION(S)	KNOWN CORRELATION(S)								
	<p><i>Describe the recommended splits and corresponding criteria.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p>	<p><i>Summarize any known correlations with the indicated features in this dataset.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate for each known correlation.)</i></p>								
	<table><tr><td>Train</td><td>62,563</td></tr><tr><td>Test</td><td>62,563</td></tr><tr><td>Validation</td><td>62,563</td></tr><tr><td>Dev</td><td>62,563</td></tr></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	Train	62,563	Test	62,563	Validation	62,563	Dev	62,563	<p><code>`field_name``,`field_name`</code></p> <p>Description: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Impact on dataset use: <Summarize here. Include visualizations, metrics, or links where necessary.></p> <p>Risks from correlation: <Summarize here. Include recommended mitigative steps if available.></p> <p>Additional Notes: <Add here></p>
Train	62,563									
Test	62,563									
Validation	62,563									
Dev	62,563									
SPLIT STATISTICS										
<p><i>Provide the sizes of each split. As appropriate, provide any descriptive statistics for features.</i></p>										

Statistic	Train	Test	Valid	Dev
Count	123456	123456	123456	123456
Descriptive Stat	123456	123456	123456	123456
Descriptive Stat	123456	123456	123456	123456
Descriptive Stat	123456	123456	123456	123456

Above: Caption for table above.

Transformations

Fill this section if any transformations were applied in the creation of your dataset.

Synopsis

TRANSFORMATION(S) APPLIED	FIELD(S) TRANSFORMED	LIBRARY(IES) AND METHOD(S) USED				
<p>Select all applicable transformations that were applied to the dataset.</p>	<p>Provide the fields in the dataset that were transformed.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each transformation type applied. Include the data types to which fields were transformed.)</p>	<p>Provide a description of the methods used to transform or process the dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each transformation type applied.)</p>				
Anomaly Detection	<Transformation Type>					
Cleaning Mismatched Values	<table><tr><th>Field Name</th><th>Source & Target</th></tr><tr><td><Field Name></td><td><Source Field: Target Field></td></tr></table>	Field Name	Source & Target	<Field Name>	<Source Field: Target Field>	Method: <Describe the transformation method here. Include links where necessary.>
Field Name	Source & Target					
<Field Name>	<Source Field: Target Field>					
Cleaning Missing Values	<table><tr><td><Field Name></td><td><Source Field: Target Field></td></tr></table>	<Field Name>	<Source Field: Target Field>	Platforms, tools, or libraries:		
<Field Name>	<Source Field: Target Field>					
Converting Data Types	<table><tr><td><Field Name></td><td><Source Field: Target Field></td></tr></table>	<Field Name>	<Source Field: Target Field>	[Platform, tool or library]: <write description here>		
<Field Name>	<Source Field: Target Field>					
Data Aggregation	<table><tr><td>...</td><td>...</td></tr></table>	[Platform, tool or library]: <write description here>		
...	...					
Dimensionality Reduction	Additional Notes: <Add here>					
Joining Input Sources						
Redaction or Anonymization						
Others (Please specify)						
	Transformation Results: <Provide results, outcomes, and actions taken because of the transformations. Include visualizations where available.>					
	Additional Notes: <Add here>					

Breakdown of Transformations

Fill out relevant rows.

CLEANING MISSING VALUE(S)	METHOD(S) USED	COMPARATIVE SUMMARY
Which fields in the data were missing values? How many?	How were missing values cleaned? What other choices were considered?	Why were missing values cleaned using this method (over others)? Provide comparative charts showing before and after missing values were cleaned.

<p><Summarize here. Include links where available.></p> <p>Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description</p>	<p><Summarize here. Include links where necessary.></p> <p>Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.></p>	<p><Summarize here. Include links, tables, visualizations where available></p> <table><tr><th>Field Name</th><th>Diff</th></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<p><Summarize here. Include links and metrics where applicable.></p> <p><Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]</p>	<p><Summarize here. Include links where available.></p>	<p><Summarize here. Include links where available.></p>								
CLEANING MISMATCHED VALUE(S)	METHOD(S) USED	COMPARATIVE SUMMARY								
<i>Which fields in the data were corrected for mismatched values?</i>	<i>How were incorrect or mismatched values cleaned? What other choices were considered?</i>	<i>Why were incorrect or mismatched values cleaned using this method (over others)? Provide a comparative analysis demonstrating before and after values were cleaned.</i>								
<p><Summarize here. Include links where available.></p> <p>Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description</p>	<p><Summarize here. Include links where available.></p>	<p><Summarize here. Include links where available.></p> <table><tr><th>Field Name</th><th>Diff</th></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									

RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								
ANOMALIES	METHOD(S) USED	COMPARATIVE SUMMARY								
<i>How many anomalies or outliers were detected? If at all, how were detected anomalies or outliers handled? Why or why not?</i>	<i>What methods were used to detect anomalies or outliers?</i>	<i>Provide a comparative analysis demonstrating before and after anomaly handling measures.</i>								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where available.>	<Summarize here. Include links where available.> <table><tr><th>Field Name</th><th>Diff</th></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></table> Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								
AGGREGATION	METHOD(S) USED	COMPARATIVE SUMMARY								

<i>Which fields in the dataset were aggregated?</i>	<i>What methods were used to aggregate the data? Include the aggregating operator. What other choices were considered?</i>	<i>Why was the data aggregated using this method (over others)? Provide comparative charts that demonstrate the choices of aggregators.</i>								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.>	<Summarize here. Include links, tables, visualizations where available.> <table><thead><tr><th>Field Name</th><th>Diff</th></tr></thead><tbody><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></tbody></table> Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								
DIMENSIONALITY REDUCTION	METHOD(S) USED	COMPARATIVE SUMMARY								
<i>How many original features were collected and how many dimensions were reduced?</i>	<i>What methods were used to reduce the dimensionality of the data? What other choices were considered?</i>	<i>Why were features reduced using this method (over others)? Provide comparative charts showing before and after dimensionality reduction processes.</i>								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.>	<Summarize here. Include links, tables, visualizations where available> <table><thead><tr><th>Field Name</th><th>Diff</th></tr></thead><tbody><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></tbody></table> Above: <Provide a caption for the above table or visualization.>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									

		Additional Notes: <Add here>								
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								
JOINING INPUT SOURCES	METHOD(S) USED	COMPARATIVE SUMMARY								
<i>What were the distinct input sources that were joined?</i>	<i>What are the shared columns of fields used to join these sources?</i>	<i>Why were features joined using this method over others?</i> <i>Provide comparative charts showing before and after dimensionality reduction processes.</i>								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.>	<Summarize here. Include links, tables, visualizations where available> <table><tr><th>Field Name</th><th>Diff</th></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></table> Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								

<Risk Type>: [Description + Mitigations]										
REDACTION OR ANONYMIZATION	METHOD(S) USED	COMPARATIVE SUMMARY								
Which features were redacted or anonymized?	What methods were used to redact or anonymize data?	Why was data redacted or anonymized using this method over others? Provide comparative charts showing before and after redaction or anonymization process.								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.>	<Summarize here. Include links, tables, visualizations where available> <table><thead><tr><th>Field Name</th><th>Diff</th></tr></thead><tbody><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr><tr><td>...</td><td>...</td></tr></tbody></table> Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									
...	...									
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS								
What risks were introduced because of this transformation? Which risks were mitigated?	What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?	What additional considerations were made?								
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>								
OTHERS (PLEASE SPECIFY)	METHOD(S) USED	COMPARATIVE SUMMARY								
What was done? Which features or fields were affected?	What methods were used?	Why was this method used over others? Provide comparative charts showing before and after this transformation.								
<Summarize here. Include links where available.> Field Name: Count or Description Field Name: Count or Description Field Name: Count or Description	<Summarize here. Include links where necessary.> Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.>	<Summarize here. Include links, tables, visualizations where available> <table><thead><tr><th>Field Name</th><th>Diff</th></tr></thead><tbody><tr><td><Field Name></td><td><Before: After></td></tr><tr><td><Field Name></td><td><Before: After></td></tr></tbody></table>	Field Name	Diff	<Field Name>	<Before: After>	<Field Name>	<Before: After>		
Field Name	Diff									
<Field Name>	<Before: After>									
<Field Name>	<Before: After>									

	[Platform, tool or library]: <Write description here.>	<table><tr><td>...</td><td>...</td></tr></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>
...	...			
RESIDUAL & OTHER RISK(S)	HUMAN OVERSIGHT MEASURE(S)	ADDITIONAL CONSIDERATIONS		
<i>What risks were introduced because of this transformation? Which risks were mitigated?</i>	<i>What human oversight measures, including additional testing, investigations and approvals were taken due to this transformation?</i>	<i>What additional considerations were made?</i>		
<Summarize here. Include links and metrics where applicable.> <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations] <Risk Type>: [Description + Mitigations]	<Summarize here. Include links where available.>	<Summarize here. Include links where available.>		

Annotations & Labeling

Fill this section if any human or algorithmic annotation tasks were performed in the creation of your dataset.

ANNOTATION WORKFORCE TYPE	ANNOTATION CHARACTERISTIC(S)	ANNOTATION DESCRIPTION(S)
<p>Select all applicable annotation workforce types or methods used to annotate the dataset:</p>	<p>Describe relevant characteristics of annotations as indicated. For quality metrics, consider including accuracy, consensus accuracy, IRR, XRR at the appropriate granularity (e.g. across dataset, by annotator, by annotation, etc.).</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each annotation type.)</p>	<p>Provide descriptions of the annotations applied to the dataset. Include links and indicate platforms, tools or libraries used wherever possible.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each annotation type.)</p>
<p>Annotation Target in Data</p> <p>Machine-Generated Annotations</p> <p>Human Annotations (Expert)</p> <p>Human Annotations (Non-Expert)</p> <p>Human Annotations (Employees)</p> <p>Human Annotations (Contractors)</p> <p>Human Annotations (Crowdsourcing)</p> <p>Human Annotations (Outsourced / Managed) Teams</p> <p>Unlabeled</p> <p>Others (Please specify)</p>	<p>(Annotation Type)</p> <p>Number of unique annotations 123456789</p> <p>Total number of annotations 123456789</p> <p>Avg. Annotations per example 123456789</p> <p>Number of annotators per example 123456789</p> <p>[Quality metric per granularity] 123456789</p> <p>[Quality metric per granularity] 123456789</p> <p>[Quality metric per granularity] 123456789</p> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	<p>(Annotation Type)</p> <p>Description: <Description of annotations (labels, ratings) produced. Include how this was created or authored.></p> <p>Link: <Relevant URL link.></p> <p>Platforms, tools, or libraries: [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.> [Platform, tool or library]: <Write description here.></p> <p>Additional Notes: <Add here></p>
	ANNOTATION DISTRIBUTION(S)	ANNOTATION TASK(S)

	<p>Provide a distribution of annotations for each annotation or class of annotations using the format below.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each annotation type.)</p>	<p>Summarize each task type associated with annotations in the dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each task type.)</p>										
	<p>(Annotation Type)</p> <table><tr><td>Annotations (or Class) (20%)</td><td>12345</td></tr><tr><td>Annotations (or Class) (20%)</td><td>12345</td></tr><tr><td>Annotations (or Class) (20%)</td><td>12345</td></tr><tr><td>Annotations (or Class) (20%)</td><td>12345</td></tr><tr><td>Annotations (or Class) (20%)</td><td>12345</td></tr></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	Annotations (or Class) (20%)	12345	Annotations (or Class) (20%)	12345	Annotations (or Class) (20%)	12345	Annotations (or Class) (20%)	12345	Annotations (or Class) (20%)	12345	<p>(Task Type)</p> <p>Task description: <Summarize here. Include links if available.></p> <p>Task instructions: <Summarize here. Include links if available.></p> <p>Methods Used: <Summarize here. Include links if available.></p> <p>Inter-rater adjudication policy: <Summarize here. Include links if available.></p> <p>Golden Questions: <Summarize here. Include links if available.></p> <p>Additional Notes: <Add here></p>
Annotations (or Class) (20%)	12345											
Annotations (or Class) (20%)	12345											
Annotations (or Class) (20%)	12345											
Annotations (or Class) (20%)	12345											
Annotations (or Class) (20%)	12345											
<h3>Human Annotators</h3>												
<p>Fill this section if human annotators were used.</p>												
	<p>ANNOTATOR DESCRIPTION(S)</p>	<p>ANNOTATOR TASK(S)</p>										
	<p>Provide a brief description for each annotator pool performing the human annotation task.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each annotation type.)</p>	<p>Provide a brief description for each annotator pool performing the human annotation task.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each annotation type.)</p>										
	<p>(Annotation Type)</p> <p>Task type. <Summarize here. Include links if available></p> <p>Number of unique annotators: <Summarize here. Include links if available.></p> <p>Expertise of annotators: <Summarize here. Include links if available.></p>	<p>(Task Type)</p> <p>Task description: <Summarize here. Include links if available.></p> <p>Task instructions: <Summarize here. Include links if available.></p> <p>Methods Used: <Summarize here. Include links if available.></p>										

	<p>Description of annotators: <Summarize here. Include links if available.></p> <p>Language distribution of annotators: <Summarize here. Include links if available.></p> <p>Geographic distribution of annotators: <Summarize here. Include links if available.></p> <p>Summary of annotation instructions: <Summarize here. Include links if available.></p> <p>Summary of gold questions: <Summarize here. Include links if available.></p> <p>Annotation platforms: <Summarize here. Include links if available.></p> <p>Additional Notes: <Add here></p>	<p>Inter-rater adjudication policy: <Summarize here. Include links if available.></p> <p>Golden Questions: <Summarize here. Include links if available.></p> <p>Additional Notes: <Add here></p>
LANGUAGE(S)	LOCATION(S)	GENDER(S)
<p><i>Provide annotator distributions for each annotation type.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate and complete the following for each annotation type.)</i></p>	<p><i>Provide annotator distributions for each annotation type.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate and complete the following for each annotation type.)</i></p>	<p><i>Provide annotator distributions for each annotation type.</i></p> <p><i>Use additional notes to capture any other relevant information or considerations.</i></p> <p><i>(Usage Note: Duplicate and complete the following for each annotation type.)</i></p>
<p>(Annotation Type)</p> <p><Language> [Percentage %] <Language> [Percentage %] <Language> [Percentage %]</p> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	<p>(Annotation Type)</p> <p><Location> [Percentage %] <Location> [Percentage %] <Location> [Percentage %]</p> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	<p>(Annotation Type)</p> <p><Gender> [Percentage %] <Gender> [Percentage %] <Gender> [Percentage %]</p> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>

Validation Types		
Fill this section if the data in the dataset was validated during or after the creation of your dataset.		
METHOD(S)	BREAKDOWN(S)	DESCRIPTION(S)
Select all applicable :	<i>Provide a description of the fields and data points that were validated.</i>	<i>Provide a description of the methods used to validate the dataset.</i>
	<i>Use additional notes to capture any other relevant information or considerations.</i>	<i>Use additional notes to capture any other relevant information or considerations.</i>
	<i>(Usage Note: Duplicate and complete the following for each validator type.)</i>	<i>(Usage Note: Duplicate and complete the following for each validator type.)</i>
Data Type Validation	(Validation Type)	(Validation Type)
Range and Constraint Validation	Number of Data Points Validated: 12345	Method: <Describe the validation method here. Include links where necessary.>
Code/cross-reference Validation	Fields Validated:	Platforms, tools, or libraries:
	Field 123456 [Count if available]	[Platform, tool or library]: <Write description here.>
	Field 123456 [Count if available]	
	Field 123456 [Count if available]	
Structured Validation	Above: <Provide a caption for the above table or visualization.>	[Platform, tool or library]: <Write description here.>
Consistency Validation		Validation Results: <Provide results, outcomes, and actions taken because of the validation. Include visualizations where available.>
Not Validated		Additional Notes: <Add here>
Others (Please Specify)		
Description of Human Validators		
Fill this section if the dataset was validated using human validators		
	CHARACTERISTIC(S)	DESCRIPTION(S)
	<i>Provide characteristics of the validator pool(s). Use additional notes to capture any other relevant information or considerations.</i>	<i>Provide a brief description of the validator pool(s). Use additional notes to capture any other relevant information or considerations.</i>
	<i>(Usage Note: Duplicate and complete the following for each validator type.)</i>	<i>(Usage Note: Duplicate and complete the following for each validator type.)</i>
	(Validation Type)	(Validation Type)
	Unique validators 12345	Validator description: <Summarize here. Include links if available.>
	# of examples per validator 123456	
	Average cost/task/ validator \$\$\$	Training provided: <Summarize here. Include links if available.>
	Training provided Y/N	
	Expertise required Y/N	Validator Selection Criteria: <Summarize here. Include links if available.>

		Training provided: <Summarize here. Include links if available.> Additional Notes: <Add here>
LANGUAGE(S)	LOCATION(S)	GENDER(S)
<i>Provide validator distributions for each validation type. Use additional notes to capture any other relevant information or considerations. (Usage Note: Duplicate and complete the following for each annotation type.)</i>		
(Validation Type) <Language> [Percentage %] <Language> [Percentage %] <Language> [Percentage %] Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	(Validation Type) <Location> [Percentage %] <Location> [Percentage %] <Location> [Percentage %] Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	(Validation Type) <Gender> [Percentage %] <Gender> [Percentage %] <Gender> [Percentage %] Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>

Sampling Methods

Fill out the following block if your dataset employs any sampling methods.

METHOD(S) USED	CHARACTERISTIC(S)	SAMPLING CRITERIA
	<p>Provide characteristics of each sampling method used.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each sampling method used.)</p>	<p>Describe the criteria used to sample data from upstream sources.</p> <p>Use additional notes to capture any other relevant information or considerations.</p>
Select all applicable methods used in the creation of this dataset:		
Cluster Sampling	(Sampling Type) Upstream Source [Write here]	<ul style="list-style-type: none">• Sampling method: <Summarize here. Include links where applicable.>• Sampling method: <Summarize here. Include links where applicable.>• Sampling method: <Summarize here. Include links where applicable.>
Haphazard Sampling	Total data sampled 123m	
Multi-stage Sampling	Sample size 123	
Random Sampling	Threshold applied 123k units at property	
Retrospective Sampling	Sampling Rate 123	
Stratified Sampling	Sample Mean 123	
Systematic Sampling	Sample Std. Dev. 123	
Weighted Sampling	Sampling Distribution 123	
Unknown	Sampling Variation 123	
Unsampled	Sample Statistic 123	
Others (Please specify)	Above: <Provide a caption for the above table or visualization.> Additional Notes: <Add here>	

Known Applications & Benchmarks

Fill out the following section if your dataset was primarily created for use in AI or ML system(s)

ML APPLICATION(S)	EVALUATION RESULT(S)	EVALUATION PROCESS(ES)								
<p>Provide a list of key ML tasks that the dataset has been used for. Usage Note: Use comma-separated keywords.</p>	<p>Provide the evaluation results from models that this dataset has been used in.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each model.)</p>	<p>Provide a description of the evaluation process for the model's overall performance or the determination of how the dataset contributes to the model's performance.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each model and method used.)</p>								
<p>For example: Classification, Regression, Object Detection</p>	<p>(Model Name)</p> <p>Model Card: [Link to Full Model Card] Evaluation Results</p> <table><tbody><tr><td>Accuracy</td><td>123 (params)</td></tr><tr><td>Precision</td><td>123 (params)</td></tr><tr><td>Recall</td><td>123 (params)</td></tr><tr><td>Performance metric</td><td>123 (params)</td></tr></tbody></table> <p>Above: <Provide a caption for the above table or visualization.></p> <p>Additional Notes: <Add here></p>	Accuracy	123 (params)	Precision	123 (params)	Recall	123 (params)	Performance metric	123 (params)	<p>(Model Name)</p> <p>[Method used]: <Summarize here. Include links where available.></p> <ul style="list-style-type: none">• Process: <Summarize here. Include links, diagrams, visualizations, and tables as relevant.>• Factors: <Summarize here. Include links, diagrams, visualizations, and tables as relevant.>• Considerations: <Summarize here. Include links, diagrams, visualizations, and tables as relevant.>• Results: <Summarize here. Include links, diagrams, visualizations, and tables as relevant.> <p>Additional Notes: <Add here></p>
Accuracy	123 (params)									
Precision	123 (params)									
Recall	123 (params)									
Performance metric	123 (params)									
	<p>DESCRIPTION(S) AND STATISTIC(S)</p>	<p>EXPECTED PERFORMANCE AND KNOWN CAVEATS</p>								
<p>Duplicate this row as necessary for each model type</p>	<p>Provide a description of the model(s) and task(s) that this dataset has been used in.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each model.)</p>	<p>Provide a description of the expected performance and known caveats of the models for this dataset.</p> <p>Use additional notes to capture any other relevant information or considerations.</p> <p>(Usage Note: Duplicate and complete the following for each model.)</p>								
	<p>(Model Name)</p> <p>Model Card: [Link to Full Model Card]</p>	<p>(Model Name)</p> <p>Expected Performance: <Summarize here. Include links where available.></p>								

	Model Description: <Summarize here. Include links where applicable.>		Known Caveats: <Summarize here. Include links, diagrams, visualizations, and tables as relevant.>
	Model Size	123 (params)	
	Model Weights	123 (params)	
	Model Layers	123 (params)	
	Latency	123 (params)	
Additional Notes: <Add here>			Additional Notes: <Add here>

Terms of Art

Concepts and Definitions referenced in this Data Card

Use this space to include the expansions and definitions of any acronyms, concepts, or terms of art used across the Data Card. Use standard definitions where possible. Include the source of the definition where indicated. If you are using an interpretation, adaptation, or modification of the standard definition for the purposes of your Data Card or dataset, include your interpretation as well.

[TERM OF ART]	[TERM OF ART]	[TERM OF ART]
Definition: <Write here>	Definition: <Write here>	Definition: <Write here>
Source: <Write here and share link>	Source: <Write here and share link>	Source: <Write here and share link>
Interpretation: <Write here>	Interpretation: <Write here>	Interpretation: <Write here>
[TERM OF ART]	[TERM OF ART]	[TERM OF ART]
Definition: <Write here>	Definition: <Write here>	Definition: <Write here>
Source: <Write here and share link>	Source: <Write here and share link>	Source: <Write here and share link>
Interpretation: <Write here>	Interpretation: <Write here>	Interpretation: <Write here>
[TERM OF ART]	[TERM OF ART]	[TERM OF ART]
Definition: <Write here>	Definition: <Write here>	Definition: <Write here>
Source: <Write here and share link>	Source: <Write here and share link>	Source: <Write here and share link>
Interpretation: <Write here>	Interpretation: <Write here>	Interpretation: <Write here>

Reflections on Data

Use this space to include any additional information about the dataset that has not been captured by the Data Card. For example, does the dataset contain data that might be offensive, insulting, threatening, or might otherwise cause anxiety? If so, please contact the appropriate parties to mitigate any risks.

[Title]	<Write notes here>
[Title]	<Write notes here>
[Title]	<Write notes here>



The [Data Cards Playbook ↗](#) by Google Research is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.

You are free to share and adapt this work under the [appropriate license terms ↗](#).