

<h1>SugarCrepe</h1>	<p>SugarCrepe is a benchmark for faithful vision-language compositionality evaluation. Given an image, a model is required to select the <i>positive</i> text that correctly describes the image, against another <i>hard negative</i> text distractor that differs from the positive text only by small compositional changes.</p>
<p>DATASET LINK</p> <p>Dataset: github.com/RAIVNLab/sugar-crepe</p>	<p>DATA CARD AUTHOR(S)</p> <p>Cheng-Yu Hsieh*, University of Washington</p> <p>Jieyu Zhang*, University of Washington</p> <p>Zixian Ma, University of Washington</p> <p>Aniruddha Kembhavi, Allen Institute for Artificial Intelligence</p> <p>Ranjay Krishna, University of Washington</p>

Dataset Overview		
DATA SUBJECT(S)	DATASET SNAPSHOT	CONTENT DESCRIPTION
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	2.7 MB
	Number of Instances	12264
	Number of Fields	3
	Labeled Classes	N/A
	Number of Labels	N/A
	Average Labels Per Instance	N/A
	Algorithmic Labels	N/A
	Human Labels	N/A
	Other Characteristics	N/A
Sensitivity of Data		
SENSITIVITY TYPE(S)	FIELD(S) WITH SENSITIVE DATA	SECURITY AND PRIVACY HANDLING

User Content User Metadata User Activity Data Identifiable Data S/PII Business Data Employee Data Pseudonymous Data Anonymous Data Health Data Children's Data None Others (Please specify)	N/A	N/A
RISK TYPE(S)	SUPPLEMENTAL LINK(S)	RISK(S) AND MITIGATION(S)
Direct Risk Indirect Risk Residual Risk No Known Risks Others (Please Specify)	N/A	N/A
Dataset Version and Maintenance		
MAINTENANCE STATUS	VERSION DETAILS	MAINTENANCE PLAN
Regularly Updated (New versions of the dataset have been or will continue to be made available.)	Current Version: 1.0 Last Updated: 06/2023 Release Date: 06/2023	We are committed and have resources to maintain the dataset. Versioning: N/A Updates: Dataset may be updated to address technical issues. Errors: We actively monitor issues at github.com/RAIVNLab/sugar-crepe

<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>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Deprecated (This dataset is obsolete or is no longer being maintained.)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: We actively monitor issues at github.com/RAIVNLab/sugar-crepe</p>
	NEXT PLANNED UPDATE(S)	EXPECTED CHANGE(S)
	N/A	N/A

Example of Data Points														
PRIMARY DATA MODALITY	SAMPLING OF DATA POINTS	DATA FIELDS												
Image Data Text Data Tabular Data Audio Data Video Data Time Series Graph Data Geospatial Data Multimodal (Image and Text Data) Unknown Others (Please specify)	[Typical Data Point Link] github.com/RAIVNLab/sugar-crepe/tree/main/data	<table> <tr> <th>Field Name</th><th>Field Value</th><th>Description</th></tr> <tr> <td>filename</td><td>string</td><td>The id to an image</td></tr> <tr> <td>caption</td><td>string</td><td>Positive text correctly describing the image</td></tr> <tr> <td>negative_caption</td><td>string</td><td>Hard negative text incorrectly describing the image</td></tr> </table>	Field Name	Field Value	Description	filename	string	The id to an image	caption	string	Positive text correctly describing the image	negative_caption	string	Hard negative text incorrectly describing the image
Field Name	Field Value	Description												
filename	string	The id to an image												
caption	string	Positive text correctly describing the image												
negative_caption	string	Hard negative text incorrectly describing the image												
TYPICAL DATA POINT		ATYPICAL DATA POINT												
<pre>{ "filename": "000000289393.jpg", "caption": "Several toy animals - a bull, giraffe, deer and parakeet.", "negative_caption": "Several toy animals - a bull, giraffe, snake and parakeet." }</pre>		N/A												

Motivations & Intentions

Motivations

PURPOSE(S)	DOMAIN(S) OF APPLICATION	MOTIVATING FACTOR(S)
Monitoring Research Production Others (Please Specify)	Machine Learning, Computer Vision, Natural Language Processing, Vision-Language Models, Compositional Understanding	<ul style="list-style-type: none">Faithfully evaluating compositionality of vision-language models.

Intended Use

DATASET USE(S)	SUITABLE USE CASE(S)	UNSUITABLE USE CASE(S)
Safe for production use Safe for research use Conditional use-some unsafe applications Only approved use Others (Please specify)	<ul style="list-style-type: none">Evaluate the compositionality of vision-language models.	N/A

LICENSING	RESEARCH AND PROBLEM SPACE(S)	CITATION GUIDELINES
We license our work using MIT License . All the source data we use is publicly released by prior work.	Evaluate the compositionality of vision-language models.	BiBTeX: <pre>``` @article{hsieh2023sugarcrape, title={SugarCrepe: Fixing Hackable Benchmarks for Vision-Language Compositionality}, author={Hsieh, Cheng-Yu and Zhang, Jieyu and Ma, Zixian and Kembhavi, Aniruddha and Krishna, Ranjay}, year={2023}, } ```</pre>