

Datasheet

We answer the questions from [26] to clarify the process of construction and accommodate transparency and accountability in our datasets.

Motivation

Q1. For what purpose was the dataset created? Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

HumaniBench was created as a *human-centric* benchmark. It fills a recognised gap in evaluating large multimodal language models (LMMs) on criteria that go beyond raw accuracy. The suite comprises seven tasks targeting *fairness, robustness, ethics, empathy, language inclusivity, understanding, and reasoning*.

Q2. Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?

The benchmark was conceived and led at the **Vector Institute for Artificial Intelligence** (Toronto, Canada). Additional contributions came from a collaborator at the University of Central Florida.

Q3. Who funded the creation of the dataset?

Development was funded by Vector Institute core research funds, supported in part by the Province of Ontario, CIFAR, and Vector’s corporate sponsors.

Q4. Other comments.

None.

Composition, Collection Process, Pre-processing / Cleaning / Labeling

Overview. The dataset contains **32,536** image–question pairs plus auxiliary labels for Tasks T1–T3. All images are RGB JPEGs with a longest edge of ≤ 1024 px. Questions are primarily in English; Tasks T4 (Multilinguality) and T6 (Empathetic Captioning) additionally include Tamil, Spanish, and Modern Standard Arabic variants. Each task is summarised in Table ??; detailed specifications appear in Section 2

Clarification on data related to people. Some instances contain recognisable people. Every image was scraped from publicly available or Creative-Commons news sources between **July 2023 - July 2024**. No new personal information was gathered.

Known skews / biases. Because all images originate from English-language news outlets, Western cultural perspectives are over-represented. This bias should be considered when interpreting model performance.

Uses

Q1. Has the dataset been used for any tasks already? If so, please provide a description.

HumaniBench has not been employed in any published work prior to this paper. All captions, social-attribute tags, question–answer pairs, and experiments were created specifically for this release under Vector Institute research-ethics approval.

Q2. Is there a repository that links to any or all papers or systems that use the dataset? If so, please provide a link or other access point.

Artefacts are indexed on the project page: <https://vectorinstitute.github.io/HumaniBench/>.

Q3. What (other) tasks could the dataset be used for?

Fine-tuning or evaluating LMMs on fairness, bias mitigation, multilingual robustness, safety alignment, and empathy captioning; data augmentation in human-centric tasks. A curated subset of offensive or biased prompts is retained deliberately to test robustness. Practitioners should filter or mask these items before deploying derived models.

Q4. Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?

The dataset intentionally includes a limited subset of offensive, inappropriate, or biased samples to probe model robustness with respect to safety, fairness, and privacy. Users are strongly advised to review and, if necessary, filter these instances before deploying models in public-facing or production settings. This caution is reiterated in both the paper and the dataset documentation.

Q5. Are there tasks for which the dataset should not be used? If so, please provide a description.

Do not use for face recognition, surveillance, or any application that profiles individuals. As the dataset includes prompts that may elicit misinformation or offensive outputs, it should not be used in public-facing applications but only for assessing MLLM reliability during development.

Distribution

Q1. Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created? If so, please provide a description.
Yes. HumaniBench will be publicly released for non-commercial research.

Q2. How will the dataset be distributed (e.g., tarball on website, API, GitHub)?

Does the dataset have a digital object identifier (DOI)?

Annotations (SHA-256 hashes, captions, labels, splits) are hosted on HuggingFace at <https://huggingface.co/vectorinstitute/HumaniBench> and mirrored via a download script in our open-source GitHub repository. A DOI will be minted upon first public release.

Q3. When will the dataset be distributed?

Target release: **June2025**.

Q4. Licence. Data and annotations are provided under CC BY-SA 4.0.. Source code is dual-licensed under MIT.

Q5. Will the dataset be distributed under a copyright or other intellectual-property (IP) licence or terms of use (ToU)? If so, please describe them and provide a link.

All images and articles were scraped from openly available web and RSS feeds identified as either public-domain or covered by permissive Creative Commons licences. We redistribute only the derived metadata (captions, questions, answers, tags, and task splits) and hashed image IDs. The released package—including all annotations and split files—is licensed under the **Creative Commons Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)** licence⁶. No fees or additional restrictions apply; users must, however, respect any residual rights attached to the original web content when retrieving it via the provided URLs. Only derived artefacts are redistributed. Users who fetch the raw images via the supplied URLs must respect any residual rights of the original publishers.

Q6. Have any third parties imposed IP-based or other restrictions on the data associated with the instances?

None known.

Q7. Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?

None.

Q8. Any other comments?

None.

Maintenance

Q1. Who will be supporting/hosting/maintaining the dataset?

The research group from vector that developed this dataset will maintain and refine it.

Q2. How can the owner/curator/manager of the dataset be contacted (e.g., email address)?

Please contact the email address provided in the paper or post issues on the official GitHub repository.

⁶<https://creativecommons.org/licenses/by-sa/4.0/>

Q3. Is there an erratum? If so, please provide a link or other access point.

Updates to the dataset, if errors are reported, will be recorded in the release history on GitHub and on the official website.

Q4. Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)? If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub).

Yes. Necessary updates—such as label corrections, additional instances, or removals—will be performed by the maintainer team at the Vector Institute on a quarterly basis (or sooner if critical issues are reported). All changes will be announced in the GitHub changelog⁷ and mirrored on the project website.

Q5. If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances? If so, please describe these limits and explain how they will be enforced.

No. We did not collect any new images or text containing personal information; all instances originated from public or CC-licensed sources. Usage restrictions therefore follow the licences of the original datasets.

Q6. Will older versions of the dataset continue to be supported/hosted/maintained? If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.

Older versions will remain archived on Hugging Face and tagged in the GitHub release history. If a version becomes obsolete (e.g., due to significant label fixes), this status will be noted in the README and changelog.

Q7. If others want to extend, augment, or contribute to the dataset, is there a mechanism for them to do so? Will these contributions be validated/verified? Is there a process for communicating or distributing these contributions to dataset consumers?

Yes. The accompanying codebase provides a scalable toolbox that allows users to integrate new splits and evaluate them on supported models. External contributions are welcome via pull requests; submissions undergo automated sanity checks (e.g., deduplication, license verification) and manual spot review before being merged. Accepted extensions are released in subsequent version tags and announced through GitHub.

Q8. Any other comments? None.

⁷<https://github.com/VectorInstitute/HumaniBench>