Curate, Connect, Inquire: A System for Findable Accessible Interoperable and Reusable (FAIR) Human-Robot Centered Datasets

Xingru Zhou, Sadanand Modak, Yao-Cheng Chan, Zhiyun Deng, Luis Sentis, Maria Esteva*

Abstract— **The rapid growth of AI in robotics has amplified the need for high-quality, reusable datasets, particularly in human robot interaction (HRI) and AI-embedded robotics. While more robotics datasets are being created, the landscape of open data in the field is uneven. This is due to lack of curation standards and consistency in publication practices, which makes finding, understanding, accessing, and reusing existing robotics data difficult. To address these challenges, we introduce a curation and access system developed through our experience curating and publishing datasets with researchers from Texas Robotics. The system integrates a data reporting template, a domain-specific knowledge graph, and a ChatGPT-powered conversational interface that enables users to explore, compare, and access robotics datasets published in an institutional data repository. The system's evaluation demonstrated that it supports consistent and correct information about, and access to data, emphasizing the importance of curation to enhance Fairness (Findability, Accessibility, Interoperability, and Reusability) of human-centered robotics datasets. Importantly, the best practices developed in this work can inform the community how to curate and publish robotics datasets. This work directly aligns with the goals of the HCRL @ ICRA 2025 workshop and represents a step towards more human-centered access to data for embodied AI.**

I. INTRODUCTION

The rise of AI in robotics has made the need for high-quality datasets for varied training applications critical. In response, researchers are increasingly creating datasets specifically for this purpose. Derived from complex and often interdisciplinary studies using mixed research methods, these often large and multimodal datasets reflect both the robots' and the humans' perspectives, some gathered in the context of carefully designed experiments and others during observations in the physical world. While various such datasets are available online, the landscape of open data for AI in robotics remains uneven at best. To begin with, finding these datasets is not straightforward. While some are published in institutional repositories with permanent digital object identifiers (DOIs), others are hosted on websites -occasionally behind restricted access mechanisms- without assurance of their medium and long-term accessibility. Because there are no agreed-upon standards for curating robotics datasets, some are described with great detail, while others are published with scant information which limits their understanding. For datasets involving human subjects, little information about the characteristics of the participants and the study conditions can conceal reasons for data usage restrictions. Because each dataset has its own content and representation, researchers have to examine them individually to determine if they are fit for reuse in a study. Coupled with the sheer size of many of them and their complex structure navigating them and identifying their quality becomes cumbersome. Not having the possibility to inquire and compare them at once is time-consuming and hinders their reusability.

Through our experience curating and publishing HRI and AI-ready datasets for different research groups, which we publish via the Texas Robotics Dataverse[1] in the Texas Data Repository (TDR)[2], we built a system to address robotics data curation and accessibility challenges. The system leverages metadata from curated datasets available in the Texas Robotics Dataverse to enable context-aware access via natural language interaction. Based on the different data curation cases, we developed a human-robotics-specific data report template and a data model to accurately represent robotics datasets' provenance, research methodology, and technologies involved. The data model is implemented as a knowledge graph within the Texas Advanced Computing Center's (TACC)[3] open infrastructure. Datasets are curated and documented according to the recommendations in the template, and their descriptive and structural metadata is automatically harvested from the Texas Data Repository, mapped to the data model, and integrated into the knowledge graph, allowing a normalized description across the different datasets. The knowledge graph schema and metadata, the data reports, and the dataset-related publications are used to train a ChatGPT-based chatbot, allowing users to query and retrieve data using natural language. To assess the chatbot, we designed a rigorous evaluation method around four information targets. The analysis of the answers, which were rated by experts, showed that the chatbot's responses are reliable, and that they retrieve the desired data. We checked the entire system against the FAIR curation principles to assess whether data are Findable, Accessible, Interoperable, and Reusable (FAIR)[4]. Figure1 shows a schematic of the automated knowledge retrieval system for HRI datasets.

This work is aligned with the goals of the HCRL @ ICRA 2025 workshop. In particular, it addresses the challenges of data accessibility for embodied AI, and promotes robust data curation practices for human-aligned robot learning within our research community. Our system represents a step towards a FAIR human-robotics data ecosystem.

This research was funded by Good Systems – Living and Working with Robots. The Texas Advanced Computing Center provided cloud computing resources.

X. Zhou, S. Modak, Y. Chan, Z. Deng, L. Sentis, and M. Esteva are with The University of Texas at Austin, TX, USA. Emails: {xingruzhou, sadanand.modak, yaocheng.chan, zdeng, lsentis}@utexas.edu;

* Corresponding author: maria@tacc.utexas.edu

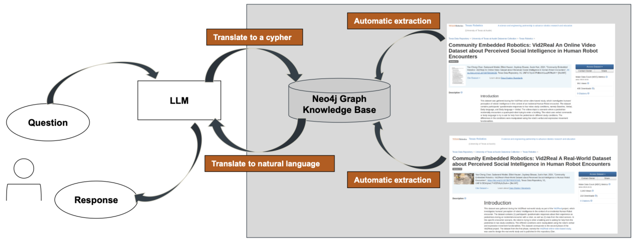


Fig. 1: Automated Knowledge Retrieval System for HRI Datasets

II. The Landscape of Large Human-Centered Robotics Datasets

While most recently, researchers have been producing a variety of robotics datasets, data curation and publication is an emerging practice in the community. In the Registry of Research Data Repositories (re3data.org), which maintains a list of international data repositories, there are no domain-specific repositories for robotics datasets, and currently no shared metadata schema or best practices have been developed to describe them. Consequently, robotics datasets are scattered and difficult to find, inconsistently organized and described, and in many cases, difficult to access. For researchers trying to find reusable robotics data in the wild, the process may involve multiple searches in different web platforms, an effort that is both time-consuming and frustrating. Even when datasets are centralized in one repository, findings can be unsatisfactory. A simple search in a general purpose digital repository like Zendo[5] using the term “Human Robot Interaction retrieves” 70 datasets, which have to be examined one by one. While some datasets, such as AFFECT-HRI[6] and HRI-CUES[7], are fairly well documented, many others are published without an explanation of what they contain or how they were collected.

P2PSTORY Dataset[8] from MIT Media Lab, UE-HRI Dataset[9] and PInSoRo Dataset[10] are datasets stored within university websites or on GitHub. These platforms are not formal data repositories, they lack permanent identifiers, and there is no guarantee of long-term sustainability. Researchers and students may move on, and websites can change. Lacking the infrastructure needed for permanent preservation and consistent citation datasets are at risk of becoming inaccessible.

Another accessibility roadblock is the size of modern robotics datasets, which may contain large numbers of heavy Rosbags and other complex image files, which are difficult to manage via a web browser. Most repositories, such as Zenodo[5], accept datasets sizes of up to 50 GB to 1 TB, and GitHub will only hold up to 100 MB[11] per project. In this work, to comply with scalable storage, long-term preservation, and ease of access, we integrate the Texas Data Repository interface with the high-performance storage backend at TACC. This approach is used to store the 4 TB CODa dataset[24].

Across the board, what is missing are shared best practices for curating human-robot datasets in ways that support long-term, cross-domain use. Without this, valuable datasets are at risk of being underused or forgotten. Our work is motivated by this gap and seeks to offer a practical solution for improving how robotics datasets are curated, accessed, and used, especially in the context of large-scale, human-centered learning. To guide the direction and components of the system, we use the FAIR principles, a set of standards that address requirements for curation and publication of datasets and for the infrastructure that hosts them.

III. Components of the FAIR Data Curation and Access System

Modern HRI experiments and real-world observations entail complex study design and cutting-edge technologies. Consequently, the derived datasets are multimodal and structurally intricate. The involvement of human subjects in the studies adds another layer in terms of assuring ethical data publication. Therefore, curation of HRI and AI-embedded robotics datasets demands a thoughtful, reproducible approach that captures the complexity of interactions transparently. To support this, we developed a system encompassing curation and access whose components we describe in the next sections. The system’s framework is shown in Figure 1.

A. Metadata Standardization, Curation Recommendations and Dataset Report

Data curation is at the system’s foundation. Curation encompasses processes such as organization, description, ethical publication, and adherence to digital preservation practices to ensure the long-term sustainability and accessibility of data[15]. Since there are no specific standards for robotics data, we achieved experience by following general curation best practices, by observing how existing datasets were publicly released[16], [17], [18], [19], [20], and through the process of curating and publishing different types of datasets for Texas Robotics research groups via the Texas Robotics Dataverse.

All curated datasets are deposited in TDR, a general-purpose institutional repository at the University of Texas Libraries that provides long-term preservation, persistent identification through DOI assignment, and centralized access. To avoid their dispersion amongst datasets from different domains, we group robotics datasets within a Texas Robotics Dataverse. We complement the system by layering robotics-specific metadata and enabling interactive, structured access to the datasets it hosts.

The baseline for describing the datasets in the repository is provided by the Dataverse Project. This is the underlying open source repository software for the TDR, which adopts, among other standards, the Data Documentation Initiative (DDI) schema[14]. Designed to describe Social Science datasets, DDI does not have elements to properly describe robotics data including collecting methods and technologies, all information needed for researchers to understand a dataset’s context and decide whether they can reuse it. The

DDI schema produces a description of the data provenance and specific social science information, useful to describe the human subjects component of an HRI dataset. As researchers deposit data, they fill in required and recommended metadata which is formatted as a JSON file that can be downloaded from the repository once the dataset is published. Using an open source repository software assures that data will be findable as the metadata is exposed to search engines and academic aggregators via standard protocols. In turn, interoperability is ensured by the use of standard metadata schemas.

So that the technical and human characteristics of the datasets are captured, we designed a data report template. This template, which we refine in iterative fashion, guides the researchers in their curation process. From the robot models and their control methods and equipment, to the experimental design or the observational methods, from the participants' roles and behavioral measures, to data post-processing methods such as segmentation and labeling, critical aspects of how datasets are collected are included in the template as descriptive elements. These will be later mapped to classes and properties in a robotics data model implemented in the Neo4J-based knowledge graph.

As we curate new datasets and encounter new elements that need to be described, we include them in a dedicated appendix section of the data report. This appendix is subcategorized to track emerging patterns across datasets. As specific element types appear more regularly, we promote them into the main body of the data report and formally incorporate them into the data model. This iterative strategy allows the system to grow and adapt while steadily moving towards broader standardization. The structured approach supports reproducibility as well as interoperability between datasets while also making them easier to understand and reuse for those outside the original research team.

After some datasets became publicly available, we received recurring feedback -through the repository- from users related to their quality and completeness. Unlike journals or conference proceedings, institutional repositories are self-publishing entities and do not have peer review in place. Thus, it is up to the researchers and curators to demonstrate a dataset's quality. Included in the data report template is a data quality statement section to record the types of quality control activities performed prior to releasing the datasets including standardized data collection (with consistent conditions and sensor calibration), annotation accuracy (verified through multi-step review and inter-rater reliability checks), and data integrity (ensured through automated and manual validation). In the case of datasets specifically created to train models, we also request that the location of the models/software is referenced preferably with DOIs, and that results of the datasets' performance or of experiments conducted for validation become part of the documentation[24], [26]. Data report guidelines addressing data quality also include using open source file formats for long-term preservation, and requesting the inclusion of data dictionaries to explain variables in tabular data.

Ethical aspects involving human subjects are carefully gauged and discussed with researchers at the point of study design and included in the report. Considering compliance with IRB decisions involving data anonymization and access restrictions, different approaches to protecting personally identifiable information can be adopted. In the case of CODa[24] recordings of incidental participants could be removed upon request, and in the Community Embedded Robotics dataset[13], participant faces were not included in the published video data. Because in many cases, facial expressions are important to capture, in[23] researchers sought informed consent for publishing such identifiable information. In this case, all but one participant were comfortable with having the recordings of their sessions released to the public without face blurring.

It has to be noted that interdisciplinary teams exhibit different opinions about privacy and data sharing. During the Robot Encounter [12] study, in which participants wearing multimodal physiological sensors to measure their stress levels shared a common space with robots, social scientists had concerns about sharing the full text of focus groups conducted after the experiment sessions, fearing that the participants identity may be recognized. Instead, roboticists considered that anonymized sensor data with ECG and EDA recordings could be openly shared pending the participants' consent. In this dataset, the sensor data is open, but only excerpts and themes resultant from the analysis of the focus groups are shared. Acknowledging the need to find a common ground for sharing human subjects' data, we identified topics that need to be reckon with by interdisciplinary teams at the design phase of an HRI study. These include, a) analyzing the degree of disclosure and sensitivity of the interview topics and potential responses, b) considering the privacy risks of all the data types that will be recorded about participants, and c) requesting participants' consent for sharing each type of data. In the data report we also require that all human subject research instruments, including surveys, questionnaires, interview protocols, and code books are published as part of the dataset to provide adequate context.

1) Scalable Organization and Access for Large Robotics Datasets: The size of a dataset is relevant to its accessibility, directly related to how the data is organized and how it can be uploaded and downloaded with ease. In terms of data organization, we work with the researchers on file organization and naming conventions that reveal the content of the dataset components. The adopted schema is explained in the data report. This is especially important to help users navigate large multimodal datasets derived from experiments with multiple sessions and tests, or involving repeated observations with different kinds of recording instruments. A data organizational schema and file naming convention (example from Vid2real real world study) is shown in Figure2. The Dataverse software allows tree views that reflect the dataset's organization, and also adding descriptive metadata to all data files, improving understanding and accessibility of the dataset.

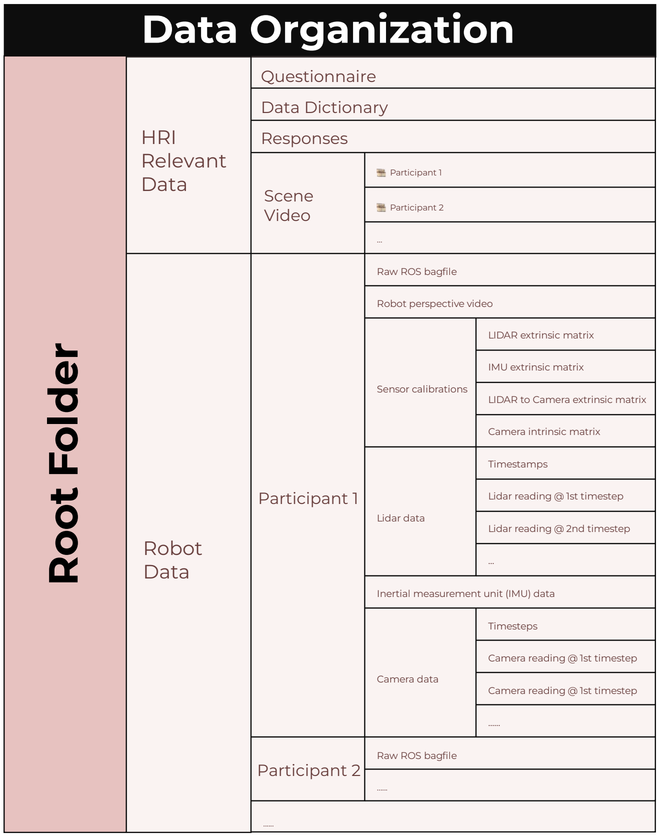


Fig. 2: Vid2real Real_World Collection Structure and Naming Convention

We mentioned that all datasets are published in TDR. However, training datasets are often bigger than the 1TB size limit storage allowed by the repository. To address this shortcoming, we partner with TACC so that data is preserved in its high performance storage resource . This hybrid approach enables permanent storage scalability while facilitating findability via the dataset's landing page and machine-readable metadata, and accessibility . In fact, part of the curation process also entails developing scripts to facilitate automated download of large datasets both from TDR and from TACC's storage resource. Prepared by the researchers in relation to how data is organized, the scripts allow access to specific sections of the datasets, an important accessibility feature for large and complex datasets.

B. A Uniform Data Model for Robotics

Through the process of curation and following the workflows narrated by researchers about their studies and how they collect and process data, we created a hierarchal data model as an abstract representation. The model defines a set of core classes and properties as metadata that reflects common and uncommon components across the different studies from which the datasets derive. Representative classes and properties include for ex. robot type - robot model -robot equipment/sensor- robot control; research method- experiment location - experiment settings - experiment session - experiment condition, etc. In the data report, each metadata field corresponds to a class or property in the data model, allowing for one-to-one mapping between contributor input and graph-ready metadata structure. Because all curated datasets conform to this shared model, the resultant metadata for each dataset is internally consistent and generalizable across datasets, making the datasets interoperable. This interoperability enables scalable integration and comparison of datasets from different sources.

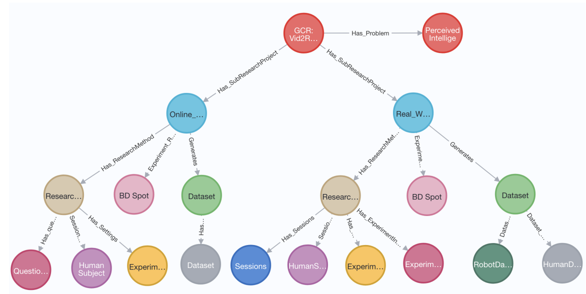


Fig. 3: Vid2real Datasets Neo4j Nodes and Relationships

C. Semantic Integration through a Knowledge Graph

Once a dataset is curated, its metadata is structured into a Neo4j-based knowledge graph according to the robotics data model. This graph converts individual metadata records into interconnected networks of nodes and relationships, as illustrated in Figure3 using as case studies the Vid2real online and real-world studies. The knowledge graph preserves curated metadata in a structured form, facilitating advanced reasoning, filtering, and integration with machine learning workflows.

Metadata extraction and knowledge graph population currently rely on structured JSON records from the Texas Robotics Dataverse datasets which employ the Data Documentation Initiative (DDI) schema. Although DDI captures general metadata elements like authors, titles, and social science details, it lacks robotics-specific as well as more detailed interdisciplinary aspects of the studies. We address this gap by integrating detailed metadata from the data report template. We employ Python-based scripts that utilize pattern matching and keyword detection (such as identifying terms like "robot", "participant", "robot model", "experiment session", "interview", "survey", "condition", etc.) to locate relevant metadata elements. These elements are systematically mapped to corresponding node types within Neo4j, ensuring precise and consistent semantic structuring. Figure4 demonstrates this mapping approach using as an example the Vid2real online study robot's metadata. This structured extraction method effectively captures both the technical and human-centered dimensions necessary for robotics research and enhances the dataset's reusability.

D. Human-Centered Access via Large Language Model

The final layer of the system connects the knowledge graph and associated files to an interactive chatbot powered by a large language model (LLM) using Retrieval-Augmented Generation (RAG). Instead of relying solely

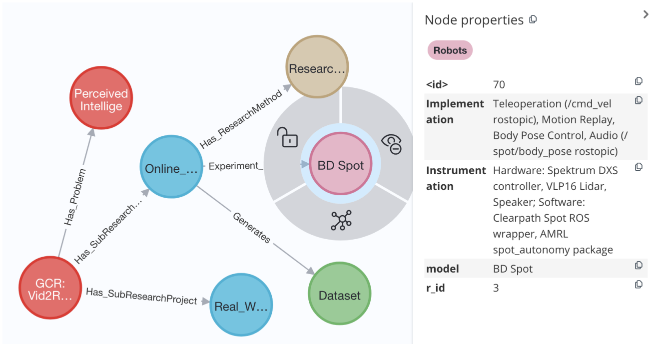


Fig. 4: Vid2real Online Study Robot Metadata Class/Node and corresponding properties in the knowledge graph.

on pre-trained knowledge, our chatbot retrieves structured metadata from the Neo4j knowledge graph, combined with supplementary materials such as related publications, IRB documentation, and selected data collection protocols and files. This comprehensive context is intended to improve the chatbot's delivery. It enables researchers to query datasets through natural language, asking questions about individual or more than one dataset, such as "What robot model was used in the Vid2Real online study?", "Which studies use a Boston Dynamic robot?" "Does the CODa dataset includes LiDAR ?", "How does the online study compare to the real-world study in terms of participant experience?" "Which research methods are involved in the online and real-world Vid2Real studies?" and receive trustworthy answers. It also allows retrieving specific files by asking questions such as "Point to all video files for session 1 in the Vid2Real real world study".

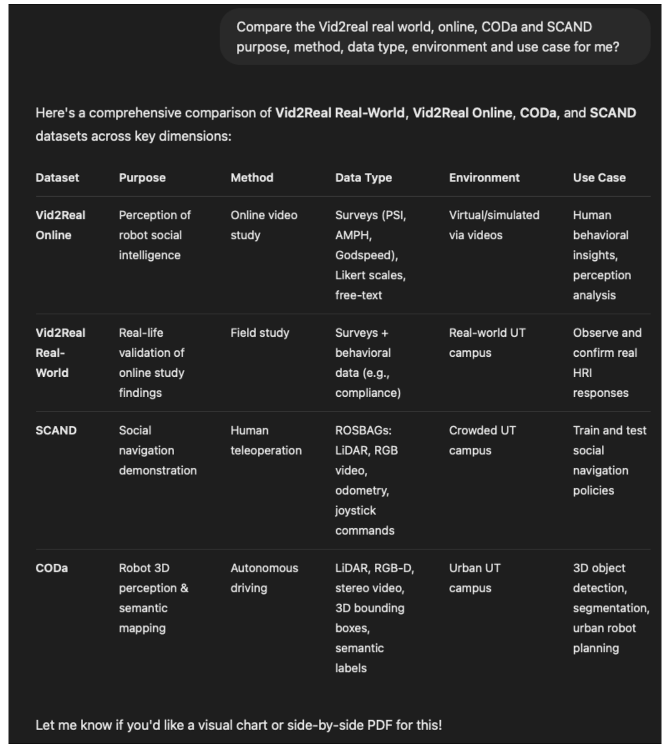
IV. Evaluation: Assessing the Performance of the Chatbot System

To evaluate how effectively our chatbot system supports interaction with robotics datasets, we conducted a human expert evaluation. Rather than measuring the model's generative performance, we focused on the chatbot's ability to deliver useful, reliable, and consistent answers across four key dimensions: information retrieval, answer stability, factual accuracy, and comparison capability.

A. Evaluation Methodology

The evaluation was conducted using four curated datasets—Vid2Real Real World [21], [23], Vid2Real Online [22], CODa [24], and SCAND [25]. Human experts familiar with robotics and data curation rated the chatbot's performance according to predefined criteria[27], [28]. Each dimension was tested using a set of ten task-specific natural language queries. Experts scored the chatbot's responses on a 0–5 scale based on the following criteria:

Information Retrieval: Ability to provide complete, relevant, and clearly structured answers to user queries regarding specific dataset content. Example queries included, "Provide me with the questionnaire link for the VID2REAL real-world study" and "List the robot behavioral conditions in the VID2REAL study."



Compare the Vid2real real world, online, CODa and SCAND purpose, method, data type, environment and use case for me?

Here's a comprehensive comparison of **Vid2Real Real-World, Vid2Real Online, CODa,** and **SCAND** datasets across key dimensions:

Dataset	Purpose	Method	Data Type	Environment	Use Case
Vid2Real Online	Perception of robot social intelligence	Online video study	Surveys (PSI, AMPH, Godspeed), Likert scales, free-text	Virtual/simulated via videos	Human behavioral insights, perception analysis
Vid2Real Real-World	Real-life validation of online study findings	Field study	Surveys + behavioral data (e.g., compliance)	Real-world UT campus	Observe and confirm real HRI responses
SCAND	Social navigation demonstration	Human teleoperation	ROSBAGs: LiDAR, RGB video, odometry, joystick commands	Crowded UT campus	Train and test social navigation policies
CODa	Robot 3D perception & semantic mapping	Autonomous driving	LiDAR, RGB-D, stereo video, 3D bounding boxes, semantic labels	Urban UT campus	3D object detection, segmentation, urban robot planning

Let me know if you'd like a visual chart or side-by-side PDF for this!

Fig. 5: Example of Evaluation: Comparison Capability

Answer Stability: Consistency of responses across semantically similar queries phrased differently. Examples of tested queries include, "What kind of robot is used in Vid2Real Real World?" versus "What type of robot was utilized in the Vid2Real real-world study?" and queries posed in multiple languages like Chinese.

Factual Accuracy: Correctness and precision of information provided based on underlying metadata. Queries assessed details such as IRB approval, sensor types, robot models, and specific experimental setups.

Comparison Capability: Ability to accurately compare two or more datasets, highlighting key differences. Example queries included, "What are the differences between Vid2Real Real World and Vid2Real Online studies regarding robot control and sensory input?" and "Which datasets use joystick-based teleoperation versus autonomous navigation?" An example comparison query and the chatbot's answer is shown in Figure5.

Scores from human reviewers were normalized using a Bayesian hierarchical model to mitigate individual rater biases. No dataset-query combination was repeated to ensure consistent and unbiased assessment across the four evaluation dimensions.

V. Results Summary

Before presenting the evaluation breakdown, we apply a Bayesian hierarchical model to adjust for individual expert preferences across criteria. Although the unadjusted scores

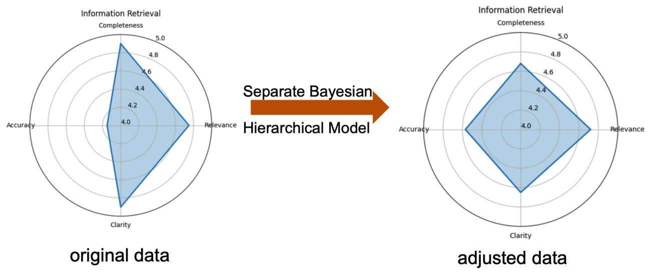


Fig. 6: Bayesian Correction in Information Retrieval

suggest highly favorable evaluations—often appearing as if the chatbot's outputs were perfectly aligned with human expectations—we use this correction to present more neutral, preference-independent scores. This approach ensures greater fairness and minimizes bias in the evaluation. The corrected scores below remain strong and reflect the chatbot's consistent performance across all aspects.

1) Information Retrieval: The chatbot achieved an average expert rating of 4.65 out of 5. In Information Retrieval, demonstrating its ability to consistently provide structured and relevant answers. Reviewers noted that responses directly referenced precise metadata elements, including links to supporting documentation, questionnaire materials, and descriptions of experimental components. Since expert preferences varied most in this aspect, we use it to illustrate how our Bayesian hierarchical model corrects for individual bias, as shown in Figure 6.

2) Answer Stability: In this dimension, the chatbot received an average score of 4.9. Responses remained consistent despite variations in query phrasing. Experts highlighted the system's robustness to linguistic variations, which significantly enhances its usability in interdisciplinary research contexts.

3) Factual Accuracy: The chatbot scored an average of 4.9 in Factual Accuracy. Human experts verified that responses accurately reflected dataset documentation, IRB statuses, robot types, and specific sensor modalities. The evaluation confirmed that the system consistently retrieved accurate, grounded information from the structured metadata and supplementary files.

4) Comparison Capability: The Comparison Capability averaged a score of 4.9. The chatbot effectively identified key differences between datasets, such as robot control methods and sensory configurations. However, it relies on precise queries that include specific dataset names. General or vague questions (e.g., "What is the robot model difference?") often yield poor results, while more targeted queries (e.g., "What is the robot model difference between CODa and SCAND?") are handled well. This highlights a key limitation: the system's ability to compare is tied to how well the user can specify their intent. Still, the chatbot successfully utilizes standardized metadata to support structured comparisons, which underscores the utility of the underlying knowledge graph. Recognizing these limitations is crucial for understanding the system's practical use and for guiding future improvements.

VI. Conclusion and Future Work

In this paper, we introduced a practical system designed to support the FAIR curation and natural language access of robotics datasets, with a focus on Human-Robot Interaction (HRI) and AI-embedded systems. Depositing datasets in an institutional repository assures findability and permanence. By making metadata interoperable through a shared data model and surfacing it via a conversational chatbot interface, we aim to make robotics datasets more accessible and reusable for a broader range of users.

Given that datasets form the foundation of most robotics research, curation and standardized metadata are a significant steps forward. Establishing these best practices within the research community could bring much-needed consistency and visibility to datasets produced in the space, offering strong motivation for their broader adoption amongst researchers that create and engage with data in robotics.

As we look ahead, our next steps include continuing to iterate on the data report template to support more detailed, domain-specific metadata. For example, by adding prompts that better reflect the needs of different HRI or embodied AI research tasks. We also plan to expand the Texas Robotics Dataverse by curating and publishing more datasets, further enriching the data landscape for human-centered robotics. In addition, we are exploring ways to add datasets from external Dataverses and repositories to the knowledge graph and chatbot. Albeit these datasets are well curated, mapping them into our data model and system will make these valuable and often underused datasets more visible, accessible, and reusable within a larger, more unified data ecosystem.

References

[1] Texas Robotics Dataverse, "Texas Robotics Dataverse," *Texas Data Repository*, 2024. [Online]. Available: https://dataverse.tdl.org/dataverse/robotics.

[2] Texas Data Repository, "Texas Data Repository," *Texas Digital Library*, 2024. [Online]. Available: https://dataverse.tdl.org.

[3] Texas Advanced Computing Center, "Texas Advanced Computing Center," *The University of Texas at Austin*, 2024. [Online]. Available: https://tacc.utexas.edu.

[4] M. D. Wilkinson *et al.*, "The FAIR Guiding Principles for scientific data management and stewardship," *Scientific Data*, vol. 3, p. 160018, Mar. 2016, doi: 10.1038/sdata.2016.18.

[5] Zendo, "Zendo: A Dataset Sharing Platform for Scientific Research," Accessed Apr. 4, 2025. [Online]. Available: https://zenodo.org

[6] J. S. Heinisch, J. Kirchhoff, P. Busch, *et al.*, "Physiological data for affective computing in HRI with anthropomorphic service robots: the AFFECT-HRI data set," *Sci. Data*, vol. 11, p. 333, 2024, doi: 10.1038/s41597-024-03128-z.

[7] B. Irfan, J. Miniota, S. Thunberg, E. Lagerstedt, S. Kuoppamäki, G. Skantze, and A. Pereira, "Human-Robot Interaction Conversational User Enjoyment Scale (HRI CUES)," arXiv preprint arXiv:2405.01354, 2024, doi: 10.48550/arXiv.2405.01354.

[8] MIT Media Lab, "P2PSTORY: Dataset of Children as Storytellers and Listeners in Peer-to-Peer Interactions," Accessed Apr. 6, 2025. [Online]. Available: https://www.media.mit.edu/projects/p2pstory/overview/

[9] A. Ben-Youssef, C. Clavel, and S. Essid, "UE-HRI: A Dataset for the Study of User Engagement in Spontaneous Human-Robot Interactions," Accessed Apr. 6, 2025. [Online]. Available: https://adasp.telecom-paris.fr/resources/2017-05-18-ue-hri/

[10] S. Lemaignan, C. E. R. Edmunds, E. Senft, and T. Belpaeme, "The PInSoRo Dataset," Accessed Apr. 6, 2025. [Online]. Available: https://freeplay-sandbox.github.io

[11] GitHub, "About large files on GitHub," *GitHub Docs*, 2024. [Online]. Available: https://docs.github.com/en/repositories/working-with-files/managing-large-files/about-large-files-on-github.

[12] R. Gupta, H. Shin, E. Norman, Z. Deng, M. Esteva, N. Lu, K. K. Stephens, and L. Sentis, "Community Embedded Robotics: A Multimodal Dataset on Perceived Safety during Indoor Mobile Robot Encounters," *Texas Data Repository*, 2024, Version 2. [Online]. Available: https://doi.org/10.18738/T8/FT9VYS.

[13] R. Gupta, E. Norman, X. Zhou, M. Esteva, K. K. Stephens, and L. Sentis, "Community Embedded Robotics: A Dataset to Study Perceived Social Intelligence and Safety During Unexpected Encounters with Quadrupedal Robots," *Texas Data Repository*, 2025, Version 1. [Online]. Available: https://doi.org/10.18738/T8/IYJES1.

[14] DDI Alliance, "Data Documentation Initiative (DDI) Metadata Specification," 2021. [Online]. Available: https://ddialliance.org.

[15] Digital Curation Centre, "About Digital Curation," Accessed Apr. 6, 2025. [Online]. Available: https://www.dcc.ac.uk/about/digital-curation

[16] M. Ahn *et al.*, "The Open X-Embodiment Dataset: Enabling Real-World Robot Learning at Scale," 2023. [Online]. Available: https://robotics-transformer-x.github.io/.

[17] E. Zhang *et al.*, "DROID: A Large-Scale In-the-Wild Robot Manipulation Dataset," 2023. [Online]. Available: https://droid-dataset.github.io/.

[18] S. A. Thompson *et al.*, "Verti-Wheelers Dataset: Human Teleoperation for Vertical Terrain Navigation," *ORC Dataverse*, George Mason University, 2020. [Online]. Available: https://doi.org/10.13021/orc2020/QSN50Q.

[19] A. Mandal *et al.*, "RH20T: Learning Diverse Robot Skills in One-Shot from Human Videos and Language," 2023. [Online]. Available: https://arxiv.org/abs/2307.00595.

[20] X. Choi *et al.*, "DexYCB: A Benchmark for Capturing Hand Grasping of Objects," 2021. [Online]. Available: https://arxiv.org/abs/2104.04631.

[21] E. Hauser, Y.-C. Chan, S. Modak, J. Biswas, and J. Hart, "Vid2Real HRI: Align video-based HRI study designs with real-world settings," in *Proc. 33rd IEEE Int. Conf. Robot. Hum. Interactive Commun. (RO-MAN)*, 2024.

[22] Y.-C. Chan, S. Modak, E. Hauser, J. Biswas, and J. Hart, "Community Embedded Robotics: Vid2Real—An online video dataset about perceived social intelligence in human–robot encounters (Version 1) [Data set]," *Texas Data Repository*, 2024, doi: 10.18738/T8/KAHJIB.

[23] Y.-C. Chan, S. Modak, E. Hauser, J. Biswas, and J. Hart, "Community Embedded Robotics: Vid2Real—A real-world dataset about perceived social intelligence in human–robot encounters (Draft version) [Data set]," *Texas Data Repository*, 2024, doi: 10.18738/T8/UOES4S.

[24] A. Zhang, C. Eranki, C. Zhang, R. Hong, P. Kalyani, L. Kalyanaraman, A. Gamare, A. Bagad, M. Esteva, and J. Biswas, "UT Campus Object Dataset (CODa) [Data set]," *Texas Data Repository*, 2023, doi: 10.18738/T8/BBOQMV.

[25] H. Karnan, A. Nair, X. Xiao, G. Warnell, S. Pirk, A. Toshev, J. Hart, J. Biswas, and P. Stone, "Socially Compliant Navigation Dataset (SCAND) [Data set]," *Texas Data Repository*, 2022, doi: 10.18738/T8/0PRYRH.

[26] W. B. Knox, S. Hatgis-Kessell, S. Booth, S. Niekum, P. Stone, and A. Allievi, "Reproduction Data for: Models of Human Preference for Learning Reward Functions," *Texas Data Repository*, 2023, Version 2. [Online]. Available: https://doi.org/10.18738/T8/S4WTWR.

[27] J. Wu, Z. Wu, R. Li, H. Qin, and G. Wang, "Effective Bug Detection in Graph Database Engines: An LLM-based Approach," *arXiv preprint arXiv:2402.00292 [cs.DB]*, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2402.00292.

[28] S. Di Bartolomeo, G. Severi, V. Schetinger, and C. Dunne, "Ask and You Shall Receive (a Graph Drawing): Testing ChatGPT's Potential to Apply Graph Layout Algorithms," *arXiv preprint arXiv:2303.08819 [cs.HC]*, 2023. [Online]. Available: https://doi.org/10.48550/arXiv.2303.08819.