STANDING Together

Recommendations for diversity, inclusivity, and generalisability in artificial intelligence health technologies and health datasets.

Published 30th October 2023.

The potential for Artificial Intelligence (AI) to benefit health must be balanced against the risks posed by algorithmic bias and harms. These technologies may work better for some groups and worse for others, causing or worsening health inequity. **To ensure the future of AI-enabled healthcare is inclusive and equitable**, the STANDING Together collaboration has developed international consensus-based recommendations to highlight and mitigate potential harms caused by bias in data and algorithms.

The recommendations across the following pages are intended to tackle inequitable performance of AI health technologies across the development lifecycle.

STANDING Together: Standards for data Diversity, Inclusivity, and Generalisability.

An extended scientific paper on these recommendations, including detailed explanatory text for each item, will be published in due course.

For further context about how to interpret these recommendations visit: www.datadiversity.org/recommendations

Part 1: Recommendations for Documentation of Health Datasets

Dataset documentation gives data meaning by describing its provenance and nature. Effective documentation enables potential users of a dataset to decide if it is suitable for their purpose, based on its content, context, and limitations.

1.1a - Dataset summary

Dataset documentation should include a summary of the dataset written in plain language. This summary should state the **Data Origin**[1] and its purpose, and give a short description of the content to help data users assess whether the dataset meets their needs.

1.1b - Dataset identity and access

Dataset documentation should:

- State the dataset's identity, including a persistent identifier and information regarding date(s) of release.
- Provide information on how the data can be accessed, including permitted use, licensing arrangements and details of the data custodian(s).
- Describe adherence to principles for data use and access.

1.1c - Reasons behind dataset creation and its purpose(s)

Dataset documentation should include the reasons why this dataset was created, including any intended benefit(s), any purposes for which dataset use should be avoided, who created the dataset (including any competing interests), and who funded it.

1.1d - Data Origin

Dataset documentation should describe the **Data Origin**, why it was selected, and what individuals were told would happen to their data.

1.1e - Data sampling and aggregation from multiple sources

Dataset documentation should describe how data were sampled from the **Dataset Source(s)**[2], including an explanation of sampling strategies, their rationale, and potential impact on the composition of the dataset. If the dataset has been compiled from multiple **Dataset Sources**, dataset documentation should describe how each source was selected, and how decisions were made during data aggregation, particularly in the case of grouping populations and modification of demographic coding.

1.1f - Data shifts over time

For longitudinal datasets or datasets with multiple versions, dataset documentation should describe any known or expected changes over time relating to the population, medical practice, or how data were collected (including devices, sensors and software used), which may contribute to data shifts over time.

---

[1] **Data origin**: the original event or context in which data were generated. For instance: data generated through provision of healthcare in a hospital, data generated upon participation in a clinical trial, data generated through interaction with a consumer-facing app.

[2] **Dataset source**: Any sources of data from which a particular dataset is created. In some cases data may be obtained directly from a Data Origin (e.g. using hospital records produced during delivery of healthcare), and in other cases data may be obtained from other sources (e.g. reproducing data sourced from other discrete datasets).

1.2a - Composition of groups within the dataset

Dataset documentation should:

- Include a summary of the groups present in the dataset. The choice of which groups to describe, and the means of categorisation should be explained.
- Highlight any known missing groups within the dataset and any reason(s) for their missingness.

1.2b - Recording of Individuals' Attributes[3]

Dataset documentation should:

- Describe how and why individuals' **Attributes** are provided in the dataset, and whether this information is available at the individual or aggregate level.
- Explain whether **Attributes** have been coded, condensed, derived or modified, stating how and why this was done.
- Highlight the proportion of **Attributes** recorded as 'unknown' or 'other' and 'prefer not to say', and if possible explain the reasons why.

1.2c - Groups at risk of disparate health outcomes

Dataset documentation should:

- Include data (when available) on **Relevant Attributes**[4] relating to individuals included within the dataset. If including these data may place Individuals at risk of identification or endanger them, these data should instead be provided at aggregate level. If data on **Relevant Attributes** are missing, reasons for this should be stated.
- Highlight the presence of groups who are at risk of disparate health outcomes caused by structural or societal factors in this dataset, with consideration of both risk factors that are universal, and those that are specific to the site of data collection.

1.3a - Limitations of the dataset

Dataset documentation should identify known or expected sources of bias, error or other factors that affect the dataset as a whole, which may impact its generalisability or applicability.

1.3b - Modifications made to the data

Dataset documentation should describe whether any data items were modified from the original source, or if any of the data is synthetic, providing the rationale for doing so and any methods used.

1.3c - Missing data

Dataset documentation should describe the proportion, nature and causes of missing data (if known), particularly if there are systematic differences across groups within the dataset. Documentation should also describe if and how missing data have been handled (e.g. imputation).

[3] **Attribute**: a measured characteristic of an individual or group of individuals which may be biological, socially constructed, or a combination of both.

[4] **Relevant attributes** encompass a wide array of characteristics which have potential associations with health outcomes, including but not limited to demographic, social and economic factors. Which Attributes are Relevant Attributes depends on the context and purpose of the dataset, and may not be consistent across countries and cultures. Depending on the context, Relevant Attributes may include (but are not limited to): sex, gender, race, ethnicity, age, socioeconomic status, sexual orientation, disability, pregnancy, relationship/marriage status, religion/belief, nationality, ancestry, occupation, language(s) spoken, caste, creed, tribe.

1.3d - Known or potential bias caused or exacerbated by data acquisition and processing

Dataset documentation should:

- Describe how bias may be introduced by the acquisition and processing of data within the dataset.
- Highlight any known or potential differences in data acquired across different groups, or differences in the uncertainty of measurements between groups.
- Describe any attempts to mitigate these biases.

1.3e - Known or potential exclusion introduced by data collection

Dataset documentation should:

- Identify the context of data collection and areas where exclusion may have been introduced into the data collection process.
- Describe any attempts to mitigate these biases.

1.3f - Known or potential bias in assigned or derived Labels[5]

Dataset documentation should:

- Provide a description of any assigned or derived **Labels**, including who decided what **Labels** to include, what they were called, and how they were generated.
- Highlight **Labels** that are at high risk of bias. For example, where **Label** generation was at the discretion of individuals, where known biases in labelling behaviour have been evidenced previously, or in the use of proxy variables.
- Describe any attempts to mitigate these biases.

1.4a - Ethics and governance

Dataset documentation should:

- State which data protection laws have been adhered to, and in which jurisdiction(s) they apply.
- Describe measures taken to protect the identities of individuals.
- Describe permissions (including ethical, legal and institutional) obtained to enable dataset creation, and details of the governance of the dataset.
- Describe adherence to principles that respect data sovereignty for communities, where relevant.

1.4b - Patient and public participation

Dataset documentation should:

- Describe the role of any advisory boards and patient and public participation groups.
- Provide information on any efforts to share data and findings with those who contributed to the dataset and any feedback gathered from participants that is relevant to data interpretation.

1.4c - Bias and impact assessments

If a formal assessment of bias, fairness or societal impact has been previously conducted on the dataset, dataset documentation should include these assessments and results.

[5] **Labels**: annotations or information which provide meaning to the raw data.

Part 2: Recommendations for Use of Health Datasets

Development of AI health technologies requires the use of data. For these technologies to improve health and wellbeing, the datasets chosen must be selected carefully and used appropriately.

2.1a - Provide sufficient information about dataset(s) to allow traceability and auditability
Datasets used in the lifecycle of AI health technologies should be accompanied by documentation which conforms to **Recommendations for Dataset Documentation** (see pages 2-4), enabling auditing against these standards.

2.2a - Identify Contextualised Groups of Interest[6] in advance who may be at risk of disparate performance or harm from the AI health technology
Data Users[7] should identify **Contextualised Groups of Interest** in advance. These may be identified in various ways, including evidence appraisal and literature review, collaboration with domain experts in the **Intended Use**[8] of the AI health technology, consultation with those who have lived experience, and evidence generation and discovery through data analysis and algorithm testing.

2.2b - Justify that datasets have been used appropriately to support the Intended Use Population[9], and Intended Use of the AI health technology
The **Intended Use Population** should be appropriately represented in datasets used in an AI health technology. The **Contextualised Groups of Interest** (i.e. those who may be at risk of disparate performance or harm, see item 2.2a) should also be included where possible, and if not included this should be explicitly stated by **Data Users**. Areas of under-representation should be identified and transparently reported by **Data Users**.

2.2c Report the explicit and implicit use of Relevant Attributes during the lifecycle of the AI health technology
Data Users should report whether and how any **Relevant Attributes** were used during the lifecycle of the AI health technology, including as a **Feature**[10], proxy or **Label**.

[6] **Contextualised groups of interest**: groups identified in advance who may be at risk of disparate health outcomes when the AI health technology is used. These groups are defined by shared Relevant Attributes which have known or suspected associations with disparate health outcomes related to the intended use of an AI health technology.

[7] **Data users:** Individuals or organisations who use a dataset in the lifecycle of an AI health technology.

[8] **Intended use** (also known as Intended Purpose): The purpose for which an AI health technology may be used, as pre-specified by the manufacturer or person/organisation legally responsible.

[9] **Intended Use Population**: The population for whom an AI health technology may be used, as pre-specified by the manufacturer.

[10] **Features**: the types of variables contained within a dataset. In the context of tabular data, features may be conceptualised as the attributes, observations, or measurements within the data, representing particular values (for instance, 'height', 'weight', 'blood sugar level'). In the context of images and other non-tabular data, features may represent a part of the data with particular relevance (for instance, the edge of an anatomical structure). In some cases Features and Attributes may represent the same phenomena.

2.2d - Evaluate performance of the AI health technology for Contextualised Groups of Interest
Data Users should report performance of the AI health technology for **Contextualised Groups of Interest** identified in 2.2a, to enable comparison of performance for each group versus aggregate performance across the overall study population, and comparison of performance between different **Contextualised Groups of Interest**.

2.2e - Identify disparate performance in any additional groups outside of the pre-specified contextualised groups of interest
As well as conducting pre-specified evaluation of performance in **Contextualised Groups of Interest** (2.2a and 2.2c), **Data Users** should also evaluate performance across other groups to identify disparate performance of the AI health technology which were not previously anticipated, and may lead to harm.

2.2f Report any approaches or methods (including fairness methods) used to intentionally modify performance across groups.
Data Users should document any attempts during the lifecycle of the AI health technology which attempt to modify performance, including addressing disparate performance across groups. If applicable, explain the rationale and goals for doing so, the methods and metrics used, whether/how thresholds were set and whether these varied between groups.

2.3a - Report limitations of datasets used, and any implications on the AI health technology.
Data Users should report the limitations of datasets used, and any implications with reference to the **Intended Use** of the AI health technology. **Data Users** should investigate whether these limitations are systematically different across relevant groups, including those with **Attributes** categorised as 'unknown', 'prefer not to say', or 'other', and report differences which could result in disparate performance of the AI health technology across groups.

2.3b - Report differences between the intended purposes of the AI health technology and datasets used, including the implications of discordance.
Data Users should report the purpose(s) of datasets used (see 1.1c), and how these differ from the **Intended Use** of the AI health technology (see 2.2b). The implications of any discordance and how this affects the suitability of the dataset for its role should be stated.

2.3c - Report findings from pre-existing assessments of the AI health technology and any datasets used.
Data Users should review any available pre-existing assessments of both the AI health technology and any datasets used, and report how the findings may have implications on groups within the **Intended Use Population**, including risk of harm.

2.4a - Address uncertainties and risks with mitigation plans.
Where **Data Users** have identified uncertainty or potentially variable performance in groups, any clinical implications resulting from these findings must be clearly stated and reported as risks. The **Data User** should document strategies to monitor, manage and reduce these risks as part of the implementation of the AI health technology.

This version (1.0) published 30th October 2023.

This document is available online from:
https://www.datadiversity.org/recommendations

Any enquiries regarding this document should be sent to:
contact@datadiversity.org

Suggested citation:
The STANDING Together collaboration. Recommendations for Diversity, Inclusivity, and Generalisability in Artificial Intelligence Health Technologies and Health Datasets. [internet]. 2023. DOI: 10.5281/zenodo.10048356
Available from URL: http://www.datadiversity.org/recommendations

Funding & support:
STANDING Together (STANdards for data Diversity, INclusivity and Generalisability) is funded by the NHS AI Lab and The Health Foundation, and supported by the National Institute for Health and care Research (NIHR) as part of the Artificial Intelligence and Racial and Ethnic Inequalities in Health and Care Award (AI_HI200014).

Acknowledgements:
The STANDING Together project team wish to thank the members of the Working Group, Consensus Group, Patient and Public Involvement and Engagement subcommittee, International Advisory Group, and all other contributors to this work. These recommendations reflect the collective insights of over 350 participants from 58 countries, without whose assistance this work could not have happened. For further details of those who developed these recommendations, please visit: https://www.datadiversity.org/people