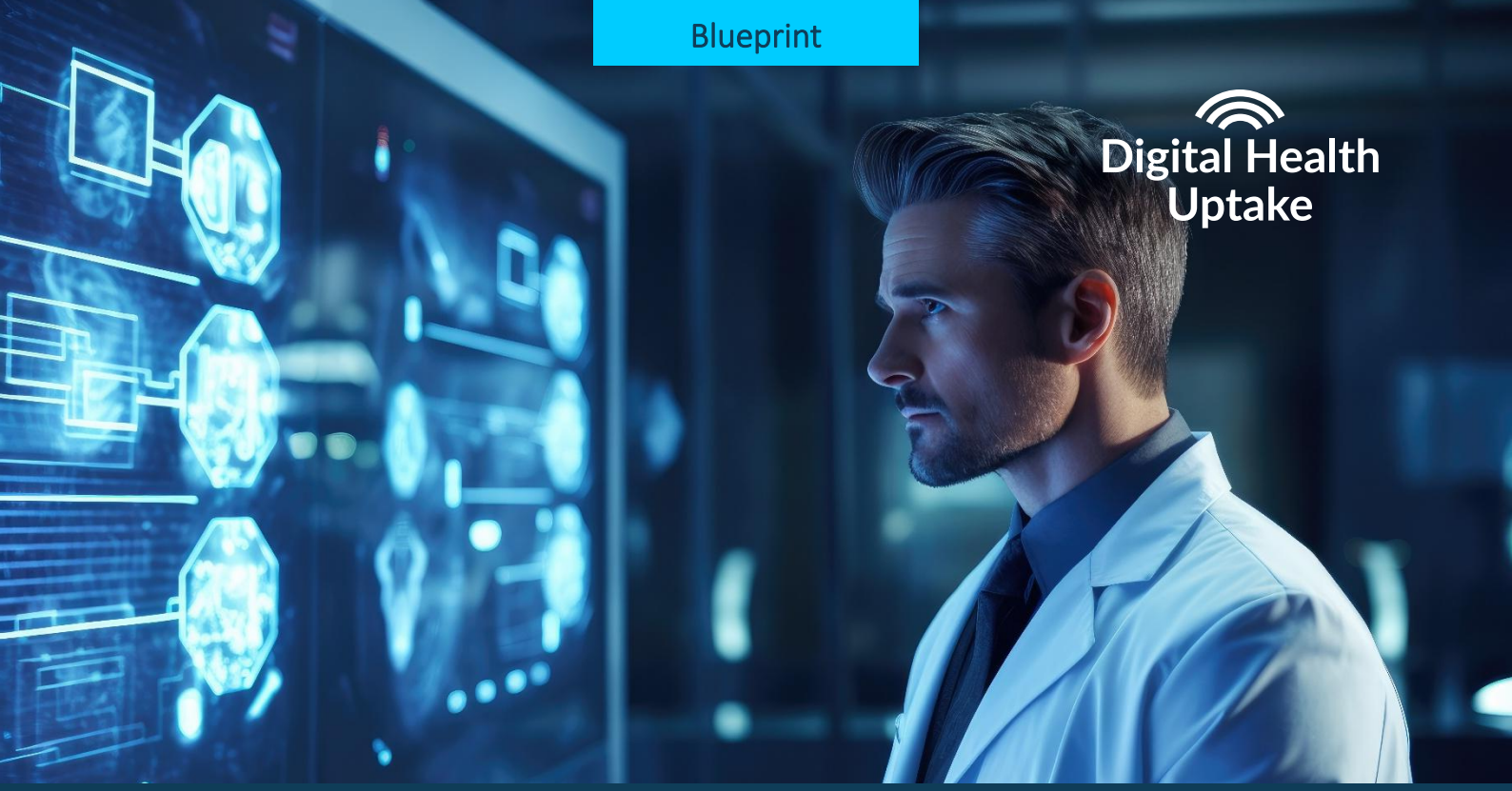
Blueprint



DHU Blueprint on the utilisation of good health data for the training of AI algorithms, in compliance with the EU AI Act

Linked to
The DHU Executive Digest:
Ensuring AI components within DHT are trustworthy, ethical and of high quality

Authors
Principal author: Dipak Kalra, i~HD
Contributors: Jens Declerck and Zoi Kolitsi, i~HD

INTENDED USE

Target DH innovation

Artificial intelligence algorithms intended for use inside AI systems to support patient care decisions within healthcare provider organisations.

Target audience

Innovators that develop AI algorithms and AI systems for use in healthcare.
Persons that assess AI systems or procure AI systems in healthcare may also find this blueprint useful to check supplied documentation on the suitability of the data used by an AI developer.

Focus of this blueprint

The use of health data sets in the training and validation of AI algorithms.

Purpose of this blueprint

To promote the adoption of good practice in the selection, quality assurance and bias assessment of health data sets, in compliance with the obligations stipulated in the EU AI Act for high risk systems. It is a requirement during AI training and validation that AI developers use data that is relevant and representative of the intended deployment patient and healthcare systems context, free from errors and biases to prevent discriminatory outcomes and to ensure fairness.

Usage

This blueprint specifies an ordered sequence of good practice steps that should be taken by the developer in order to ensure that suitable high-quality data is used for AI development. These steps should be initiated prior to obtaining and using health data sets. The decisions, choices and assessments made through the steps should be documented as part of AI explainability evidence.

BLUEPRINT DEVELOPMENT PROCESS

Primary sources of information used

- The EU AI Act
- The European ALTAI checklist
- ISO8000 (series of standards on data quality)

Synthesis

This blueprint combines information from the above with experts active in this field, including the author and contributors listed above, and goes beyond the state of the art of current evidence.
This blueprint draws on (as yet unpublished) work being undertaken by the authors in the following European Commission supported projects:
PARADISE, IDERHA, AIDAVA

Further reading

Recent initiatives that may in future develop additional data quality and bias guidance are:

- EC consultation on the Code of Practice for general-purpose Artificial Intelligence
- The Trustworthy & Responsible AI Network (TRAIN) initiative

BLUEPRINT

1 Perform a risk assessment

- Perform a comprehensive risk assessment as would be required for any high risk AI development in Europe that support patient care decisions.
- Extract and note any risk areas that relate to the use of training and validation data during AI development.
- Plan risk mitigation measures and regularly monitor these when executing the rest of this blueprint.

2 Set up a data quality process

- Appoint a data quality manager to oversee the implementation of data quality process
- Establish a data quality Standard Operating Procedure (SOP)

3 Select appropriate data

- Identify data sources that correspond well to the healthcare contexts in which it is intended for the AI system to eventually be used. This includes using data from the same country, or at least countries with equivalent populations, socio-economic characteristics, healthcare system models and pro rata corresponding healthcare budget.
- Utilise data from equivalent healthcare provider systems to the intended system use, for example from hospitals if hospital deployment is the intended eventual use.
- If data is being used from non-healthcare sources such as disease registries, claims databases, then a subset of the data sets should be used that originate from the relevant healthcare provider organisation types.
- Utilise recent enough data for which the standard of care (such as the available portfolio of investigations and treatments for the intended patients) is equivalent to the standard of care that will be prevalent when and where the AI system is used.
- Ensure through one or a combination of data sources that a sufficient sample size of relevant patients has been used. Consider statistical methods to determine the suitable sample size and document the methodologies used.
- If more than one data source is used, ensure these are consistent or knowingly and deliberately heterogeneous in order to span an appropriate diversity of contexts.

4 Select training and validation data covering appropriate patients

Formally define first the target patient profile for the intended use of the AI system. This profile might include the following characteristics:

1. Age distribution
2. Gender distribution
3. Race, ethnicity and cultures
4. Socio-economic status
5. Lifestyle factors
6. The principal condition(s) for which the AI system will be used, its name, severity, interval since original onset or diagnosis
7. Disease trajectory lifecycle points at which the AI system will be used
8. Comorbidity patterns
9. Standards of care being used e.g. prevalent clinical guideline(s)
10. Medication usage pattern

For each relevant characteristic, first find the distribution of the observed values in the relevant patient population. For example, the distribution of disease severity or the frequency distribution of treatments used for the same condition at the same severity, within that country or region

Statistically compare the value distribution within each data set to be used for AI development (training and validation) with the population norms, in order to verify the representativeness of each dataset or to discover biases that may need statistical correction before use.

Specify the care pathway point or points at which the AI system is intended to be used, and verify a sufficient number of patients and data at that care pathway point within each dataset to be used.

5 Select and extract appropriate data

Create proactively a list of the data elements that are expected to be needed in order to enable the algorithm to make appropriate novel insights and to appropriately sub-stratify the population in order to generate personalised guidance.

Examine the availability of these data elements per data source, considering the possible requirement to utilise NLP to complement the structured and coded data.

It is likely that mapping will be necessary from data source structures and terminology systems, possibly also units of measure, to the standardised data specifications which will be used by the machine learning. Document all mapping specifications and document the versions of mapping tables and terminology systems that were used during the data transformation process, which person was responsible in the organisation and the version of all tools used.

Similarly document any natural language translations that were applied in order to homogenise the language of text values across diverse data sets.

Perform spot checks on the mappings in order to verify the quality and integrity of the data transformations undertaken.

6 Assess data quality

- Define the data quality dimensions that are relevant to the data processing to be undertaken through machine learning, per data element, and the order of assessment.
- For each dimension per data element, specify the rule that will be applied and the minimum acceptability threshold.
- Undertake the data quality assessment using recognised statistical tools, and document the results of the assessment.
- Examine the overall result of data quality assessment per dataset, and determine:
  1. if one or more data sets as a whole need to be rejected on quality grounds;
  2. if some data elements need to be removed from the dataset because their individual quality is insufficient;
  3. if statistical corrections can be applied to the values of specific data elements, such as imputation of missing values.
- Carefully document all decisions and actions taken, per data set.

7 Assess representativeness and bias

- Document any statistical methods used to examine and correct for bias and which dimensions were looked at for this, including any synthetic data generation to compensate for under-represented sub populations.
- Define in advance, monitor for and document the assessments and any corrective actions that were required for algorithmic bias, evaluation bias and interactive bias.

8 Assess representativeness and bias

- Re-examine the risk assessment that was initially undertaken, relating to the use of health data.
- Confirm and document if some of the pre-planned mitigations were activated and the outcome in terms of risk mitigation measures applied and risks successfully mitigated.
- Document any other risks that arose during AI development and how they were mitigated.

9 **Document the use of data**	Ensure that all of the steps undertaken through this blueprint have been documented precisely and clearly, as documentation that may form part of an AI Act certification dossier, or may be required for Medical Device Regulation certification or for later HTA approvals.

Funded by the European Union

Digital Health Uptake (DHU) receives funding from the European Union's Digital Europe Programme under grant agreement number 101083929.