FORHUMANITY

Understanding “Data” in AI, Algorithmic and Autonomous Systems – A Concept Framework

February 2023

https://forhumanity.center/

This article will guide the reader through understanding of data in the context of Independent Audit of AI Systems and ForHumanity’s audit criteria on UK GDPR, the EU Artificial Intelligence Act and our own Risk Management Framework.

I. ABSTRACT

The world of AI, Algorithmic and Autonomous (AAA) systems is an amalgamation of many different areas of expertise, all reliant upon some form of "data", including:

- data science,
- data engineering,
- software development,
- hardware engineering,
- testing & evaluation, including accuracy and validity testing (e.g., —-software testing) and NTSB-esque safety testing (e.g., aviation safety protocols),
- ethics,
- law,
- risk management, including finance (e.g, model risk management) and residual risk disclosure (e.g., clinical drug trials).

The terms and definitions surrounding "data" are often inconsistent with each other in these varied areas of expertise. Given the significant importance of "data" (e.g, Ethical Choices, Quality considerations, Collection, Management and Governance processes) in socio-technical systems, understanding the many meanings attached to the word "data" in a holistic and precise way is necessary to evaluate the risk implications (ethics, bias, privacy, trust and cybersecurity) to humans. Such evaluation enables optimizing risk mitigations associated with "data", for…Humanity.

This concept note will document the many facets, forms, transformations and progressions of "data" through the lifecycle of AAA systems in four dimensions (hierarchy, nature, task and risk/reward profile).

CONTRIBUTORS

- Ryan Carrier, *Executive Director, ForHumanity*
- Sundar Narayanan, *ForHumanity Fellow*
- *Greg Elliott, ForHumanity Fellow*
- *Esther Chung, ForHumanity Fellow*
- *Carol Anderson, ForHumanity Contributor*
- *Sher Nicholson, ForHumanity Contributor*
- *Maud Stiernet, ForHumanity Contributor*
- *Joshua Scarpino, ForHumanity Contributor*

II. INTRODUCTION

AAA systems are socio-technical systems. Humans are embedded in the AAA System from the beginning of the process all the way to the end. Whether through the use of Personal Data or outcomes, AAA Systems run the risk of negative impacts to humans and/or our humanity. Through differentiation and precision, ForHumanity aims to integrate the law (EU Artificial Intelligence Act/GDPR) and industry practice with Independent Audit of AI Systems to ensure Governance, Oversight and Accountability of every moment and discrete occurrence of "data".

Data transits across the algorithmic lifecycle and is ever present along the way - existing both before the AAA System is designed and after the AAA system is decommissioned. It exists in many forms, frequently concurrently and thus the conundrum of taxonomy. The concurrence and multidimensional nature of data demands clarity and specificity.

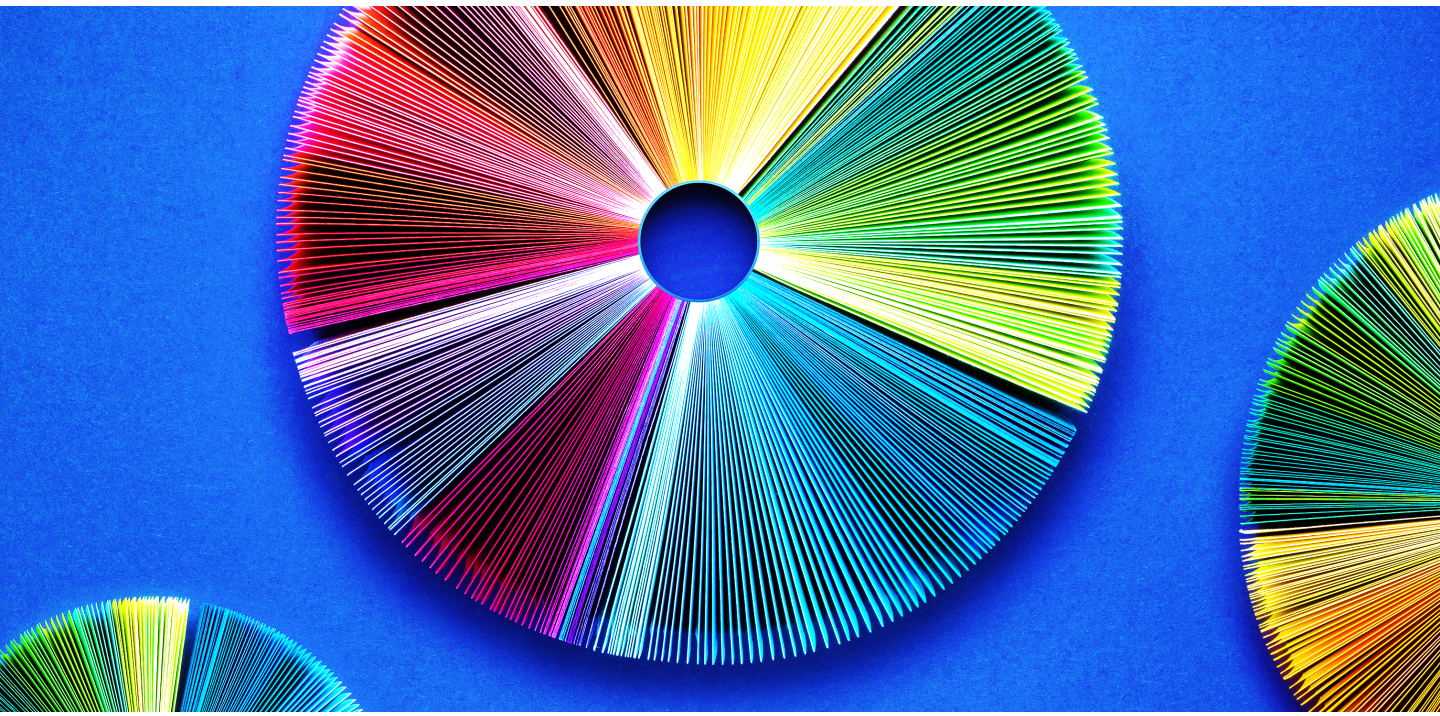
Precise understanding of Data is necessary because when misused, the potential for harms or negative outcomes for natural persons exists. When well defined and well understood, we can assess, analyze, evaluate and maximize risk mitigations associated with data across the entire lifecycle of AAA Systems.

The Concept framework establishes four dimensions of "data" identified in the graphic above and described below.

1. Data Hierarchy - from bottom to top - Data to Information to Knowledge to Wisdom

2. Nature of Data - represented by distinct shapes in the graphic
    - Dataset
    - Test Results
    - Stats & Metrics
    - Model Outcomes
    - Adverse Events
    - Metadata

3. Risk/Reward profile changed by activities during the Algorithmic Lifecycle
    - Data at Source
    - Preprocessed
    - Testing and Evaluation Data
    - Pipeline Data
    - Outcomes Data
4. Task - described at each step of the Process Flow
    - Data Collection
    - Data Labeling
    - Data Cleaning
    - Data Transformation and Reduction
    - Training and Testing/Validation split
    - Model Design
    - Model Testing and Validation
    - Model Tuning
    - Model Deployment
    - Model Integration
    - Human-in/on-the-loop
    - Model Health Fitness and Monitoring
    - Post Market insights/Feedback



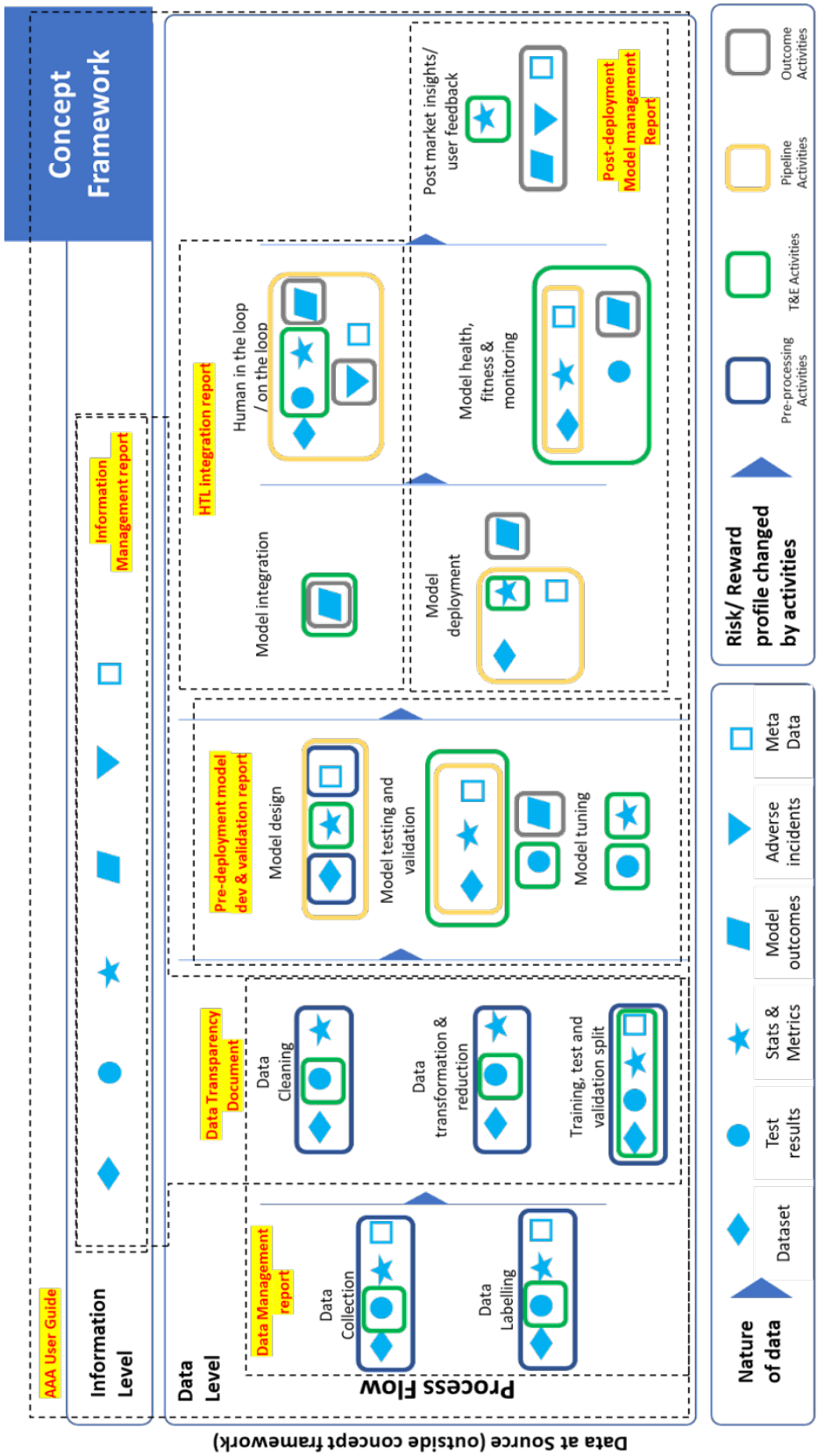


Figure 1 below will be our ever-present and evolving guide on this journey.

The graphic also provides a summation of technical documentation required under the EU Artificial Intelligence Act (as proposed) and established as audit criteria in ForHumanity's certification scheme. These are captured in Red words and defined later in the document. Technical Reports include:

- Data Management
- Data Transparency Document
- Information Management
- Pre-Deployment Model Dev and Validation
- HTL Integration
- Post Deployment Model Management

III. UNDERSTANDING THE FRAMEWORK

A. Data Hierarchy

Data Hierarchy contains four hierarchical levels: data, information, knowledge, and wisdom.

As data journeys across the AAA life cycle, many processes and procedures are applied to the data. These processes and procedures can enhance the data, reach conclusions, determine inferences, and sometimes causality may be estimated or even discovered. When data becomes useful, that is our definition of information. Information confirmed as actionable and valuable is called Knowledge, and Knowledge in the context of multi-stakeholder and optimization of all impacts is Wisdom.

Data Level - the temporal state of a collection of individual pieces of datum, where each datum is not yet known to be accurate, complete, authentic, well-described, and/or useful

RELEVANT TERMS

Data Quality - Data that is expected to be fit for purpose, representative and aligned to the Scope, Nature, Context and Purpose of the intended use as applicable to an AAA System. Data Quality is characterized as complete, accurate, categorically representative, consistent, precise collected from reasonably calibrated sensors, surveys or other tools to gather data

Information level - the temporal state of a collection of an individual piece of datum, where each datum in the collection is expected to be accurate, complete, authentic, well-described, and is expected to be useful.

RELEVANT TERMS

Information Quality - Data that is fit for purpose, representative and aligned to the Scope, Nature, Context, and Purpose of the intended use as applicable to an AAA System. Information Quality is characterized by Construct validity, Provenance, authority, and data age, including consent or other lawful basis associated initial provision of the data

PHILOSOPHICAL VIEW OF DATA HIERARCHY

Knowledge Level - the temporal state of a collection of individual pieces of datum having reached the information level and combined, examined, tested against, and in combination with each other to produce meaningful inferences, conclusions or causal relationships.

Wisdom Level - the temporal state of knowledge whereby the user of the information recognizes that knowledge alone cannot account for the consequences (intended or unintended) of applications of knowledge. Wisdom balances consequences against comprehensive health and well-being for all impacted stakeholders.

B. Nature of Data

With AAA Systems, data is everywhere. Many data types are deployed to achieve different purposes - to make a successful system. We identify six discrete types of data to accurately identify the presence and purpose of data at any moment in time during the AAA System lifecycle. All six data types are present in the life cycle at some point; however, referring to "the data" at any point in the life cycle will "muddy the waters" of understanding at best and more than likely create genuine confusion creating a greater opportunity for errors and thus negative impact to humans.

For clarity's sake, parameters, variables, hyperparameters, weights, and other elements used to establish an algorithmic calculation or process are considered Architectural Inputs and, for this concept framework, are NOT considered "data".

Dataset - a collection of datum (either raw or wrangled) intended for use in AI, algorithmic or autonomous systems (features or labels are subsets of a dataset)

- Raw – the capture of data, prior to any manipulation (e.g. cleaning, labeling, organizing) as acquired from any source in its original form
- Wrangled – a collection of data, post enhancement by cleaning, labeling, organizing

Test Results - outputs of mathematical calculations performed on a dataset or training-testing/validation dataset split designed to understand or explain characteristics of a dataset.

Metrics/Stats - thresholds, levels, other predetermined (or externally designated) measurements that are used to do any of the following:

- characterize a dataset or model performance
- determine sufficiency or insufficiency and require further action

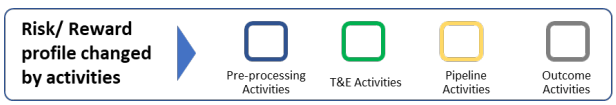
Model Outcomes - the results from an AAA System (prediction/classification)

Adverse Incidents - negative outcomes or impacts reported by a natural person resulting from AAA System and logged in an Adverse Incident Reporting System (AIRS)

Metadata - information about a datum (e.g. location, owner, date, time)

C. Risk/Reward profile changed by activities during the Algorithmic Lifecycle

Risk/Reward captures the idea that as data is manipulated during its journey through the algorithmic lifecycle - designers, developers and data scientists plan to create value (better understanding). However, ForHumanity recognizes that the creation of value through manipulation introduces risk. This distinction is identified in the framework to ensure that all users consider and understand the change and impact that may have occurred in the data.



The above image highlights the change in data's risk/reward profile as a function of activities in the algorithmic lifecycle.

Data are differentiated when new risks are introduced -

Data at Source - all data prior to determining expected usefulness based on scope/nature/context/purpose for use with the chosen model or process and outside of the scope for any audit associated with the AAA System. (technically, this data is outside of this taxonomy)

Pre-processing Activities- associated with collecting, labeling, cleaning, transforming, and reducing data before deploying the data in the model

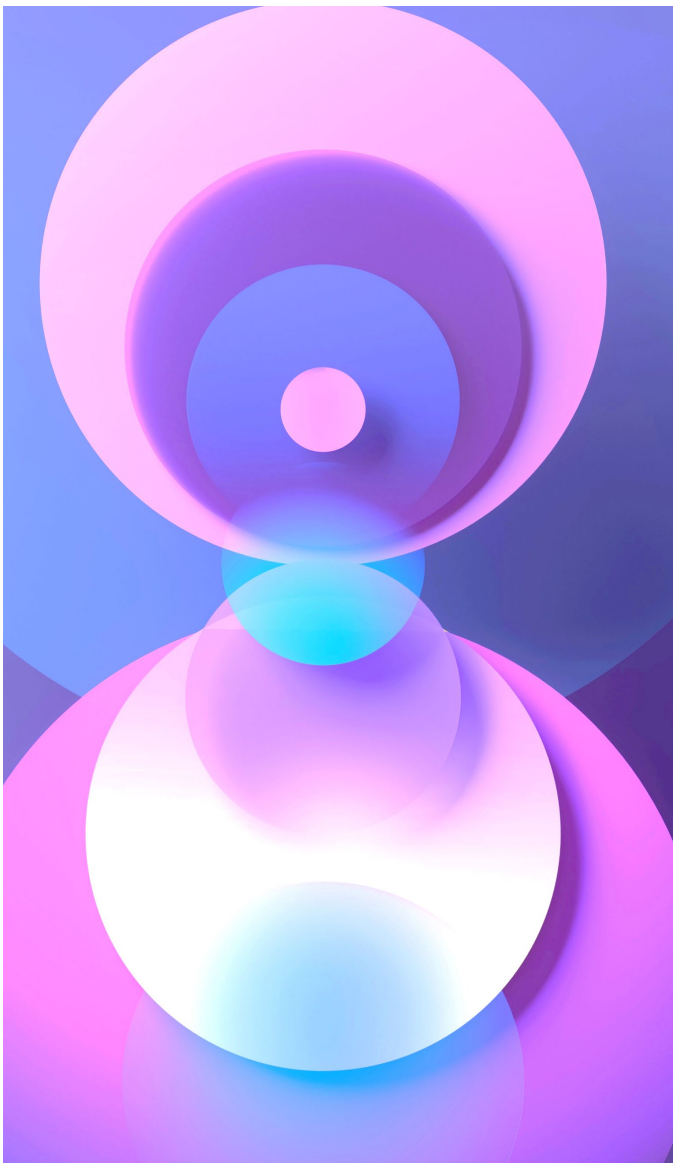
Testing and Evaluation Activities- associated with the interaction of test data, examples, and results, including metrics and thresholds associated with model development, model tuning, and monitoring for health and fitness (including ground truth)

Pipeline Activities- (includes production activities) associated with the inputs to an operational AAA System, during testing and live deployment, from external sources (including natural persons) via a predetermined collection mechanism

Outcomes Activities - associated with real-world results of a AAA System, including predictions, recommendations, post-market and continuous monitoring, and Adverse Incident Reporting Systems

D. Task

The task explains the general operation performed on the data type and classification along the AAA System life cycle. Each data task may be tailored, curated, or customized and is not a rigid classification scheme but is offered as a general explanation of the task being undertaken upon both the data type and data classification.



E. Understanding Technical Documentation

Technical documentation focused on data required by many laws and regulations worldwide. The EU AI Act places considerable responsibility on AI, Algorithmic, and Autonomous Systems providers to report or disclose data usage. These reports and documents encompass more than just data (e.g., processes and assessments). For more information on the comprehensive nature of these reports, please see ForHumanity's Certification Scheme Criteria for the EU AI Act.

The red words in the diagram describe requirements found in the EU AI Act Annex IV or ForHumanity specified audit criteria regarding data. These reports are meant to provide greater transparency and understanding of data usage in some cases to sophisticated integrators of AAA Systems and, in other cases, to provide valuable disclosure to the public.

Data Management report (EU AIA): Documenting the data requirements in terms of datasheets describing:

- the sourcing of data
- data transformations including labeling and cleaning techniques,
- nature and management of data for training/testing/validation splits

Data Transparency Document (FH): Supplement to both the Data Management Report and Information Management Report capturing all relevant steps taken by the **Ethics Committee** and the **Algorithm Risk Committee** to:

- mitigate the risk of **Bias**, to
- determine sufficient representativeness and to
- justify when, how, and why synthetic data is added into the dataset.

Information Management Report (EU AIA): Documentation of the process to determine:

- fitness for use (scope, nature, context and purpose),
- Construct Validity,
- accuracy of data in representing ground truth,
- Provenance,
- authority,
- data age,
- consent or other lawful basis associated with the initial provision of the data.

Pre-deploy Model Development and Validation Report (EU AIA): Documents the methods and steps performed for the development of the AAA system:

- the design specifications of the system , including the general logic of the AAA system,
- key design choices, rationale, and assumptions associated with it,
- training methodologies and techniques,
- the validation and testing procedures, metrics, measurement, and thresholds used for pre-deployment evaluation.

Post-deployment Model Management Report (EU AIA): **D**ocument Pipeline Quality and data input format and documents the process, metrics, measurements, and thresholds for:

- model fitness
- determining to decommission
- continuous monitoring
- post-market monitoring
- establish an Adverse Incident Reporting Systems (AIRS)

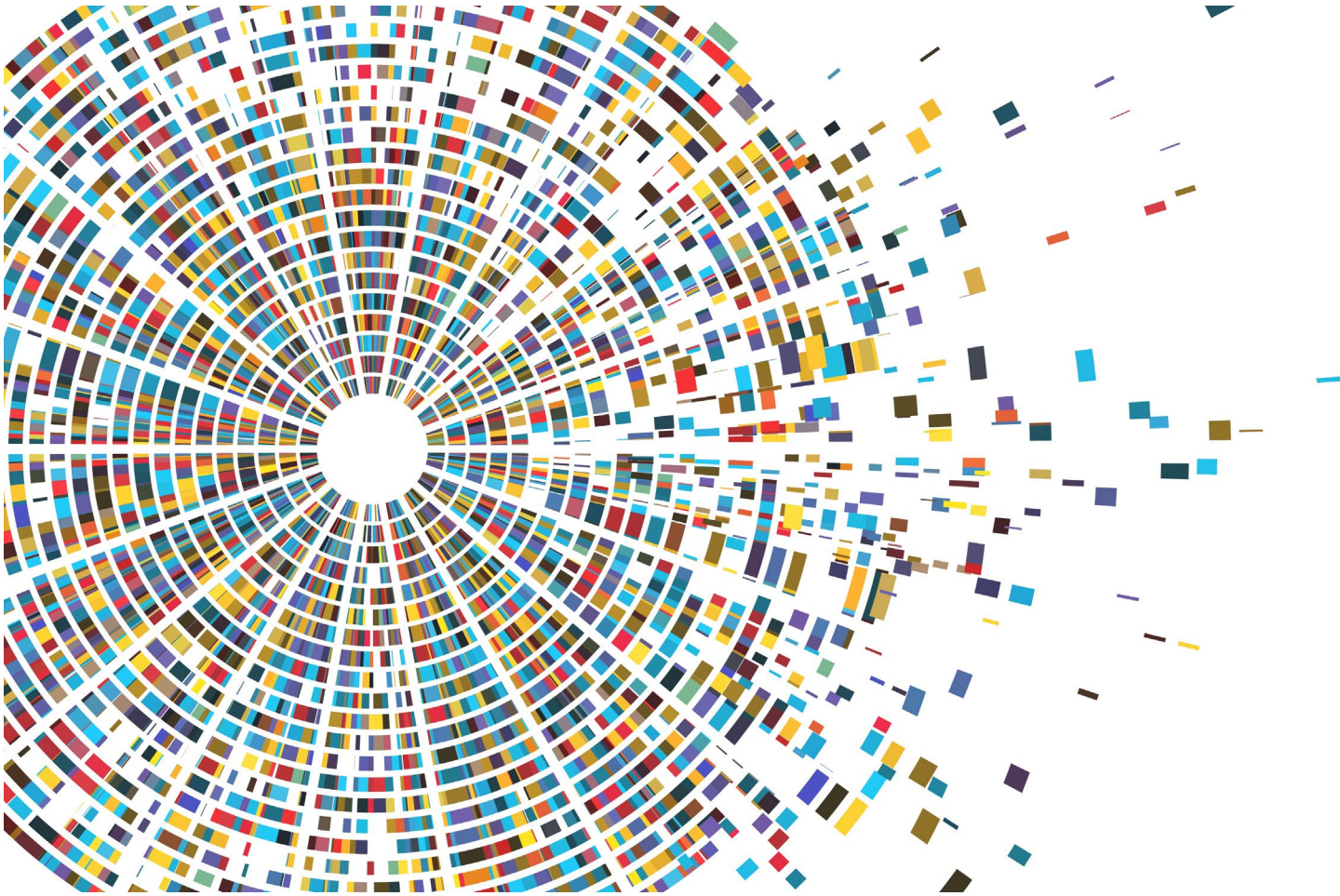
HTL and Integration Report (EU AIA): Documentation on human and system interactions, including a RACI (Responsible, Accountable, Consultation, Information), training, processes and procedures, and assessment of the human/system interface and effectiveness. Also, covers Exceptions Interpretability specifications.

AAA User Guide (EU AIA): Comprehensive guide capturing many of the elements from the aforementioned documents. It intends to enable and empower Users to successfully operate the AAA System. It is digital information that is concise, complete, correct, clear, relevant, accessible and comprehensible to Users.

It shall include guidance, and where applicable installation instructions, for monitoring, operating, and controlling the AAA system, in particular with regard to:

- its capabilities and limitations in performance.
- guidance for integration and interface between the AAA system and the User
- user-defined inputs,
- procedures to duly monitor operations
- Exceptions Interpretability interface guidance to prevent or detect anomalies, dysfunctions and unexpected performance

As AAA Systems proliferate, the risk of negative outcomes and impacts attributable to data increases. We hope this concept framework increases the precision with which designers, developers, data scientists, integrators, risk managers, decision-makers, technical and non-technical users can identify the many forms of data. Using this precision, a risk assessment should be easier at each step of the process allowing for robust analysis, evaluation, and subsequent risk treatment. Enhanced risk treatment across the lifecycle of data in AAA Systems will lead to further risk mitigation and enhanced ability to comply with the proposed EU AI Act and other regulations worldwide.



FORHUMANITY

About ForHumanity: ForHumanity (https://forhumanity.center/) is a 501(c)(3) nonprofit business dedicated to addressing the Ethics, Bias, Privacy, Trust, and Cybersecurity in artificial intelligence and autonomous systems. ForHumanity uses an open and transparent process that draws from a pool of over 1100+ international contributors to construct audit criteria, certification schemes, and educational programs for legal and compliance professionals, educators, auditors, developers, and legislators to mitigate bias, enhance ethics, protect privacy, build trust, improve cybersecurity, and drive accountability and transparency in AI and autonomous systems. ForHumanity works to make AI safe for all people and makes itself available to support government agencies and instrumentalities to manage risk associated with AI and autonomous systems.

Images: Microsoft