A **Data-Centric** Approach for **Trustworthy AI** in **Journalism**

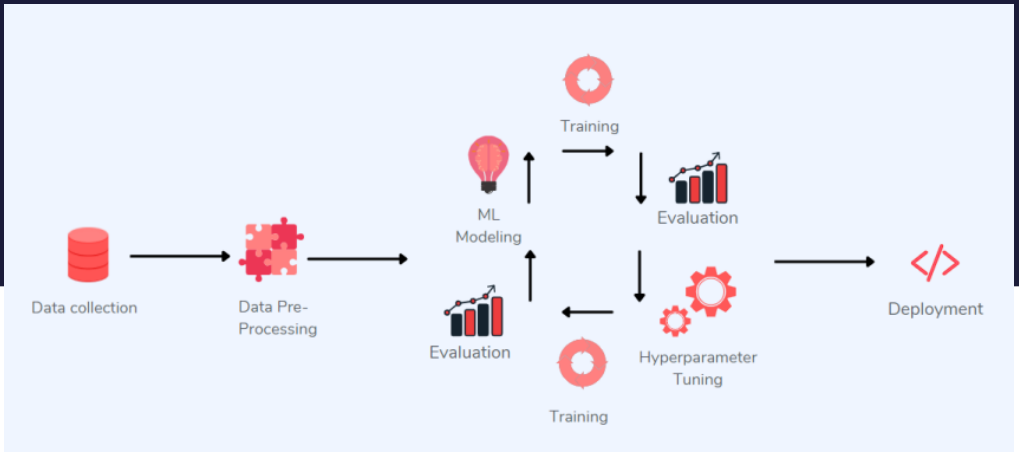
Laurence Dierickx, Carl-Gustav Lindén, Andreas L Opdahl,
Sohail Ahmed Khan, Diana Carolina Guerrero Rojas

University of Bergen

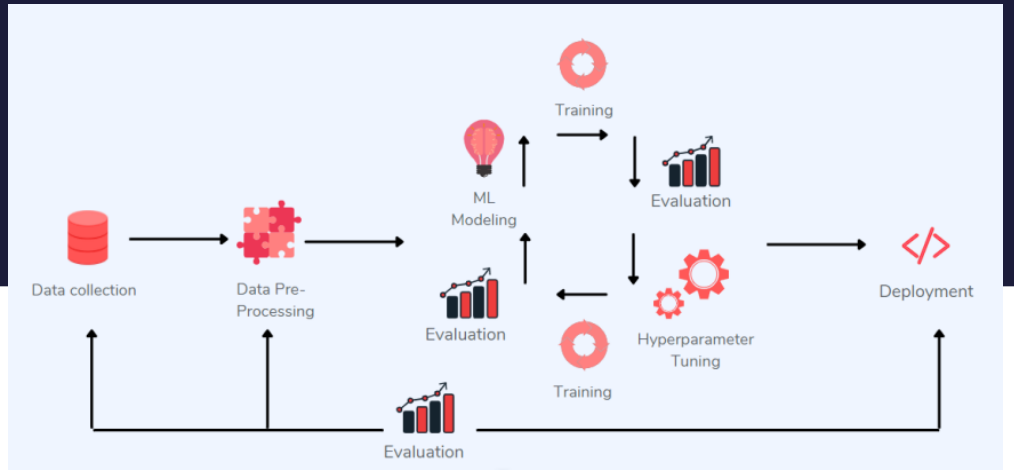
Amsterdam, September 2023

Model Centric vs Data Centric

- Big data model (as much data as possible) = volume
- Optimising the model (finding the best ([hyper]parameters) = performance
- Iterative



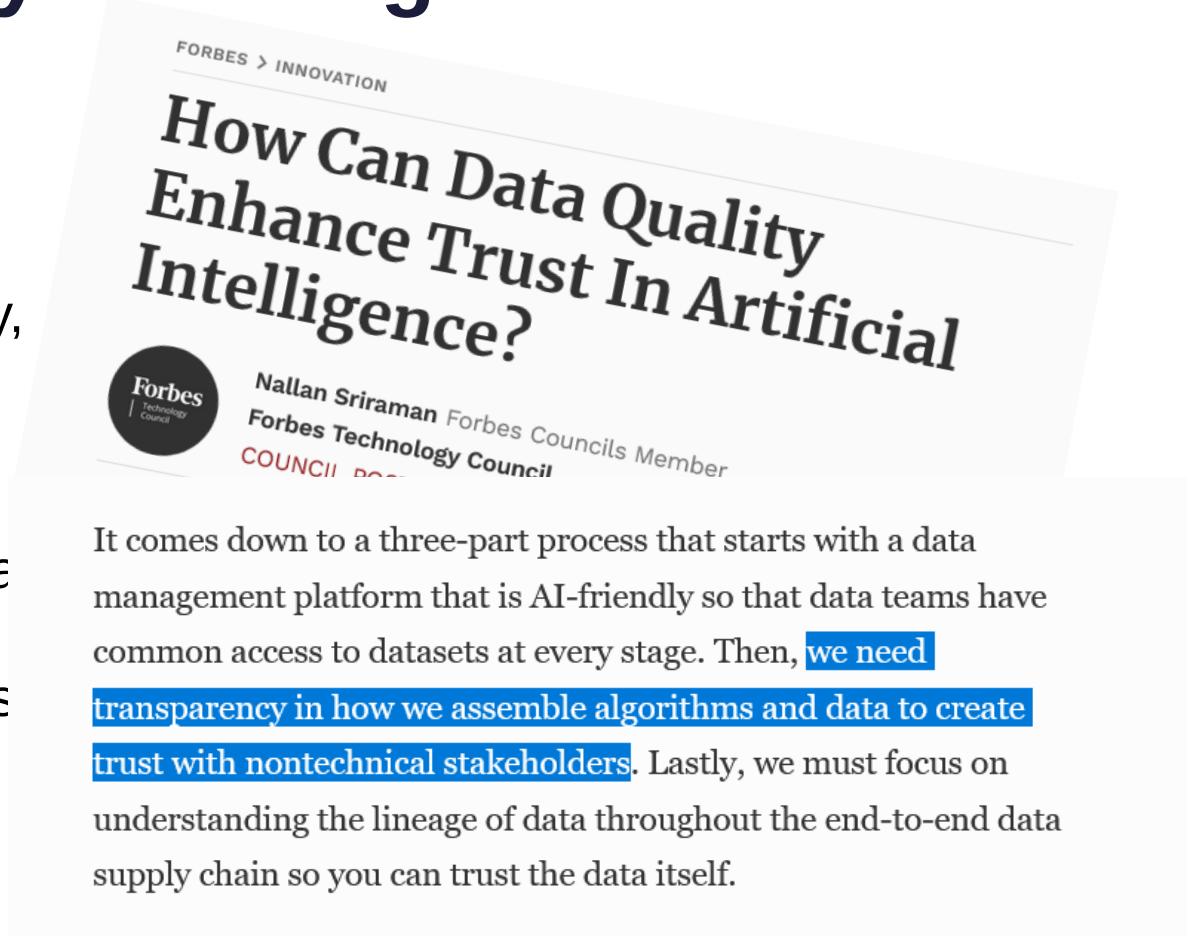
- Less data, focus on data consistency = quality
- Improving data quality = DQ tools/strategies
- Allows better customisation (fitness-for-use, expert knowledge)
- Deals better with missing values



Addressing the data quality challenges

- Data that adapts to end-uses (accuracy, relevance, understandability)
- Specific challenges in big data (believability/credibility, verifiability, consistency, interoperability)
- In ML: data quality influence the outputs (volume, completeness). Relates to data acquisition, datasets, and algorithm (as data ha impact on it)
- Prevention, detection, and correction strategies (trustworthiness, reliability)

Trusting the system is trusting the data on which it relies



FORBES > INNOVATION

How Can Data Quality Enhance Trust In Artificial Intelligence?

Forbes Technology Council

Nallan Sriraman Forbes Councils Member

Forbes Technology Council

It comes down to a three-part process that starts with a data management platform that is AI-friendly so that data teams have common access to datasets at every stage. Then, we need transparency in how we assemble algorithms and data to create trust with nontechnical stakeholders. Lastly, we must focus on understanding the lineage of data throughout the end-to-end data supply chain so you can trust the data itself.

Acquisition, pre-processing, evaluation, validation

Pre-processing challenges

Depends as much on the specificities of the data, including the types of variables as on the algorithm used

KNN: does not handle missing data well
Naïve Bayes: not ideal for large datasets with many numerical features
Decision tree: a minor change in the training data can lead to significant changes in the outcomes

Data must be available in sufficient quantity to allow training a model of sufficient quality (in terms of measures such as accuracy, precision and recall and, in a wider sense, fit for use).

Ethical data quality in journalism

In data-driven journalism: right to use and extract the data, right to privacy, transparency to foster trust towards audiences (“opening the black box”)

In AI-driven journalism: transparency on data, algorithms at work, outcomes BUT not always easy to implement AND journalists still lack data and algorithm literacy

GenAI intensified debates on ethical practices in AI journalism

Ethical codes and recommendations emphasis on human responsibility, “human-in-the-loop” approach

Foundations of the assessment framework

- Drawing on data and computer science + ethical AI in journalism + previous works
- Improving explainability & understandability (data collection, data pre-processing, data labelling)
- Integrating journalist/expert knowledge (data quality is use and context dependent)
- Applications: assessing existing datasets (reusability), developing new training datasets

Accuracy, fairness, transparency

Shanks' semiotic approach on data quality (1999): syntactic and semantic (truth, consistent structure and accurate meaning), pragmatic (fairness, usable, and useful) and social (transparency, shared knowledge)

- **Accuracy**
  Accuracy, consistency, correctness, comprehensibility ("facts as they are")
- **Fairness**
  Timeliness, completeness, accessibility, objectivity, relevance, usability
- **Transparency**
  Reliability, credibility, verifiability

Ethical	Semiotic	Dimension	Verification
Accuracy	Syntactic	Accuracy	- Level of interoperability, standardisation - Ratio accurate values/total values (a measure of erroneous data) - Uniqueness (measurement of duplicate entries and redundancies) - Encoding problems and information overload
		Consistency	- Well-defined data structure (percentage of data with consistent format and values) - Homogeneity vs heterogeneity (format, structure, values), in particular when data come from multiple sources - Unambiguous and explicit labelling
	Semantics	Correctness	- Identifying abnormal values - Identifying the causes of NULL values - Evaluation of the spelling coherence - Data documented with metadata and compliant with the metadata
		Understandability	- The extent to which data are comprehensible (feedback from the end-user)
Fairness	Pragmatic	Timeliness	- Currentness (percentage of updated data)
		Completeness	- Appropriate amount of data (ratio missing values/total values, ration NULL values/total values)
		Accessibility	- Right to use the data (access, licence, terms of use) - Level of retrievability of the data
		Objectivity	- Unbiased data (size and representativity) - Identification of human bias (annotations included)
		Relevance	- The extent to which the data are relevant for the purpose (feedback from the end-user) - Newsworthiness (evaluation of the journalistic added value of the use of the data, and of the expected impact on audiences, feedback from the end-user) - Data scarcity (measurement of the fraction of data containing relevant information)
		Usability	Fitness-for-use (to assess globally through the formal and empirical indicators of the frameworks, according to the purpose of the project = making sense of AI in a journalistic context) How automation structures and presents the data (outputs)
Transparency	Social	Reliability	- Authenticity (source) - Authority (source, annotators) - Reputation (source, annotators)
		Credibility	- Degree of the believability of the data source - Degree of the believability of the data - Degree of the believability of the annotation process and of the annotators
		Verifiability	- Verification of the source and the data - Verification of the annotation process

Benefits for AI-driven journalism

- Blending AI systems with journalistic values
- Moving from a technical to a sociotechnical approach
- Application of the fitness-for-use principle (ISO 9000)
- Fostering (a needed) interdisciplinarity
- Improving the transparency and explainability of the system: data collection, data pre-processing, data labelling
- Integration of expert knowledge (data quality is context and use dependent), requires an expertise of the application domain
- Tool for AI literacy in journalism

Paper’s outlines

1. Introduction: situating the problem, tackling data quality (GIGO) and quality in journalism

2. Data quality in machine learning: literature review, grounded in research works related to relational database + additional challenges with big data and machine learning where outcomes strongly depend on the data, considered non-technical audience but requires a bit of technicity

3. Ethical foundations for AI-driven journalism: aims to define the core ethical principles of journalism

4. Building the Accuracy-Fairness-Transparency (AFT) framework: considering 2 and 3, definition of the levels of assessment (formal and empirical indicators, aims to provide tools about the validation)

5. Conclusion: emphasises the need for interdisciplinarity / common preoccupations

Thank you!