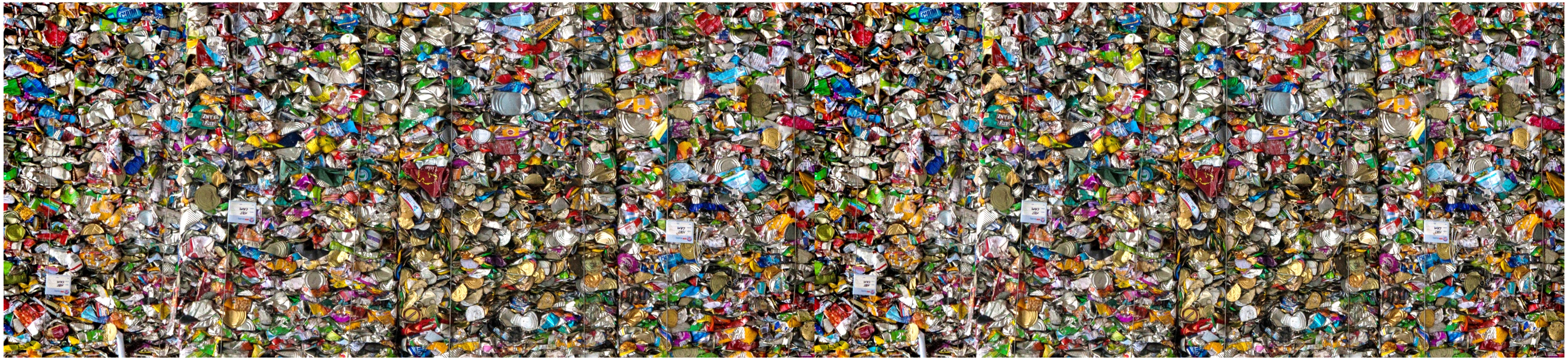

Datorios
things go
better
with
Cake
Small Mistake, Huge Difference.
Ensure Data Quality.
7
4

University of Amsterdam
Faculty of Science



Beyond Accuracy: Data Quality as the Backbone of Trustworthy AI
Waterloo University

September 22, 2025
Hazar Harmouch – h.harmouch@uva.nl
indelab.org

Short Bio

September 2015 | August 2020 | November 2023 | September 2025

PhD | Post Doc | Assistant Professor

HPI Hasso Plattner Institut

Universität Potsdam

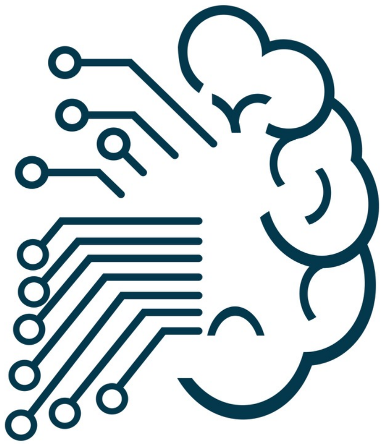
INDE lab
INtelligent Data Engineering

UNIVERSITEIT VAN AMSTERDAM

What we do

We investigate intelligent systems that support people in their work with data and information from diverse sources.

In this area, we perform applied and fundamental research informed by empirical insights into data science practice.



INDE lab

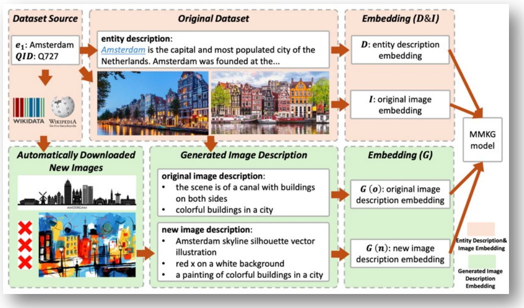
INtelligent Data Engineering

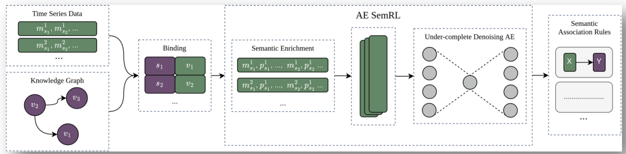
UNIVERSITY OF AMSTERDAM
Faculty of Science

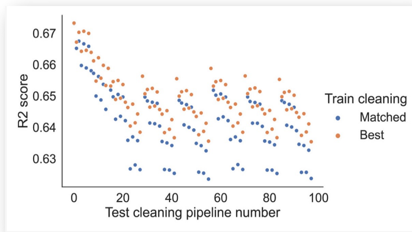
INDE lab
INtelligent Data Engineering

Research Topics at INDE lab

- **Automated Knowledge Graph Construction**
  (e.g. building KGs from multiple modalities; architectures for integrating KGs and LLMs)
- **Context Aware Data Systems**
  (e.g. rule learning & digital twins; human-data interaction; human - ai workflows)
- **Data Management for Machine Learning**
  (e.g. data quality assessment; data handling impact on ML models; data search)







University of Amsterdam
Faculty of Science

The Group - September 2025


Prof. Paul Groth


Dr. Frank Nack


Dr. Victoria Degeler


Dr. Jan-Christoph Kalo


Dr. Hazar Harmouch


Dr. Daphne Miedema


Dr. Lise Stork


Dr. Na Li


Shubha Guha


Mina Ghadimi Atigh


Dmitrii Orlov


Pengyu Zhang


Corey Harper


Danru Xu


Fina Polat


Bradley Allen


Antonis Georgakopoulos


Lucas Lageweg


Zeyu Zhang


Erkan Karabulut


Teresa Liberatore


Yichun Wang


David Jackson


Imane El Ghabi

Guests


Dr. Hartmut Koenitz


Till Döhmen


Dr. Klim Zaporojets

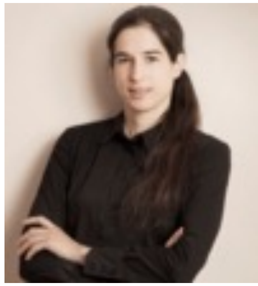

Thiviyan Thanapalasingam

INDE lab
INtelligent Data Engineering
started Nov. 5, 2018

University of Amsterdam
Faculty of Science

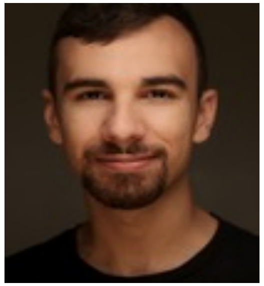
Hasso Plattner Institut

Metis Project



Lisa Ehrlinger

Divya Bhadauria



Sedir Mohammed

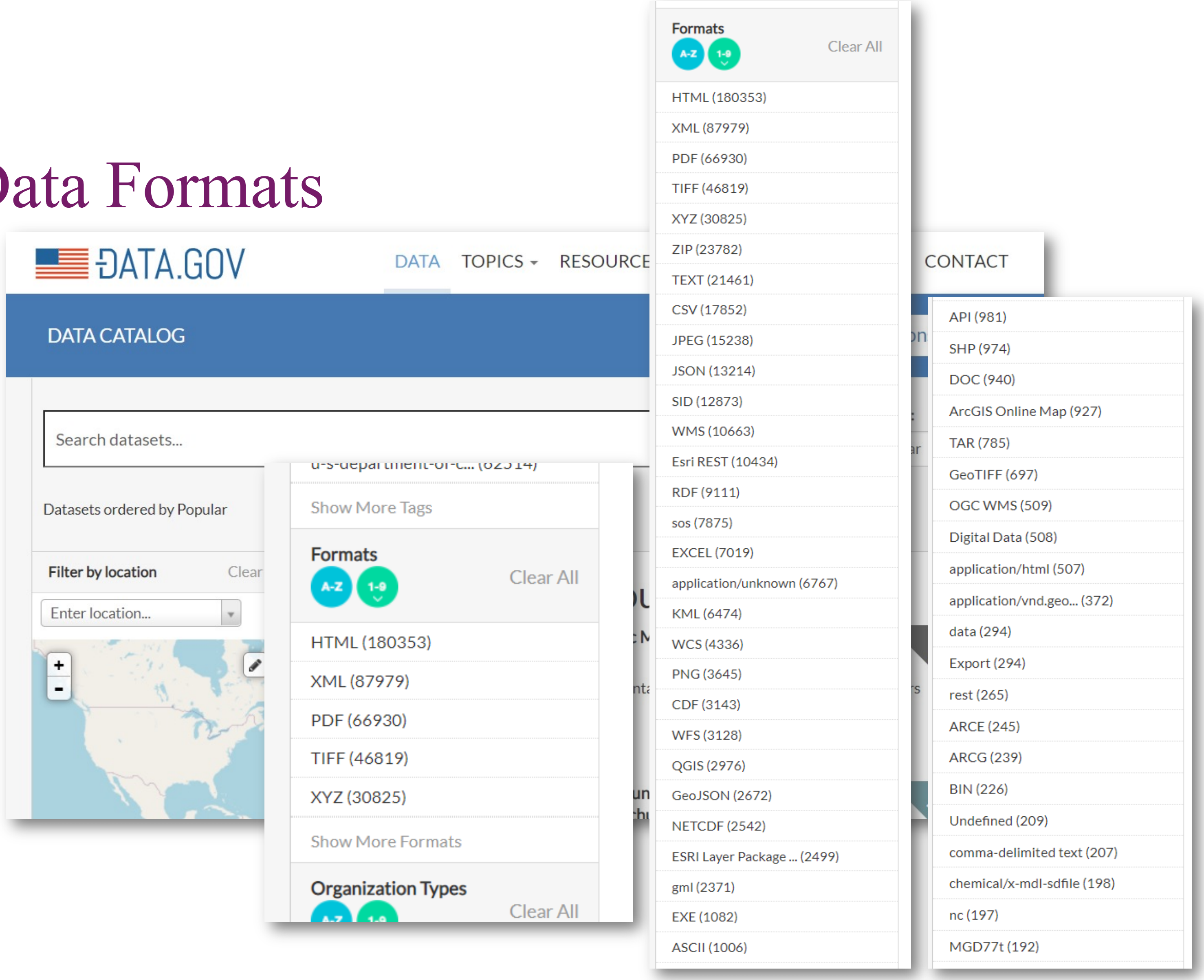
Divesh Srivastava

Felix Naumann

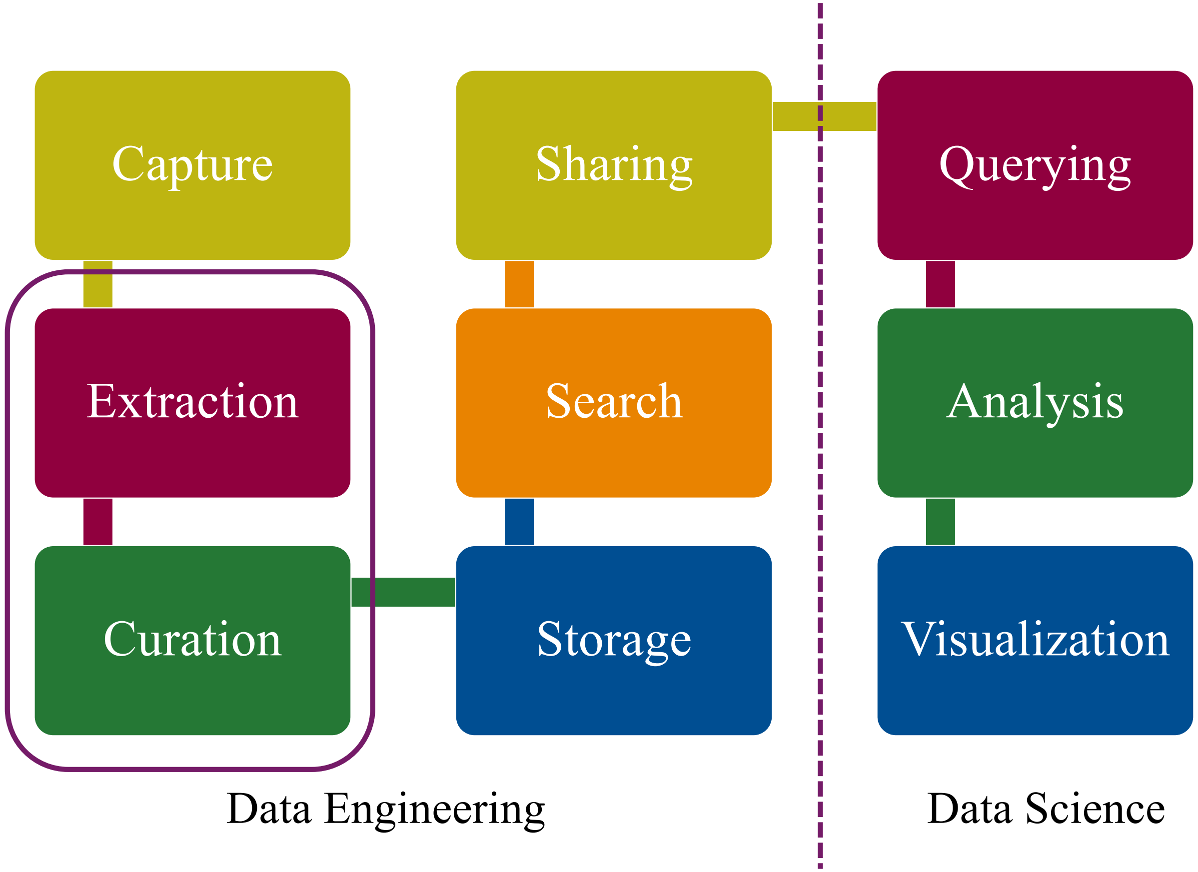
https://hpi.de/naumann/projects/data-integration-data-quality-and-data-cleansing/metis.html

Data Sources – Data Formats

- Data lakes
- Web tables
- Open (government) data
- Instrumented processes
- Sensor data
- Experimental output
- Database exports
- Excel

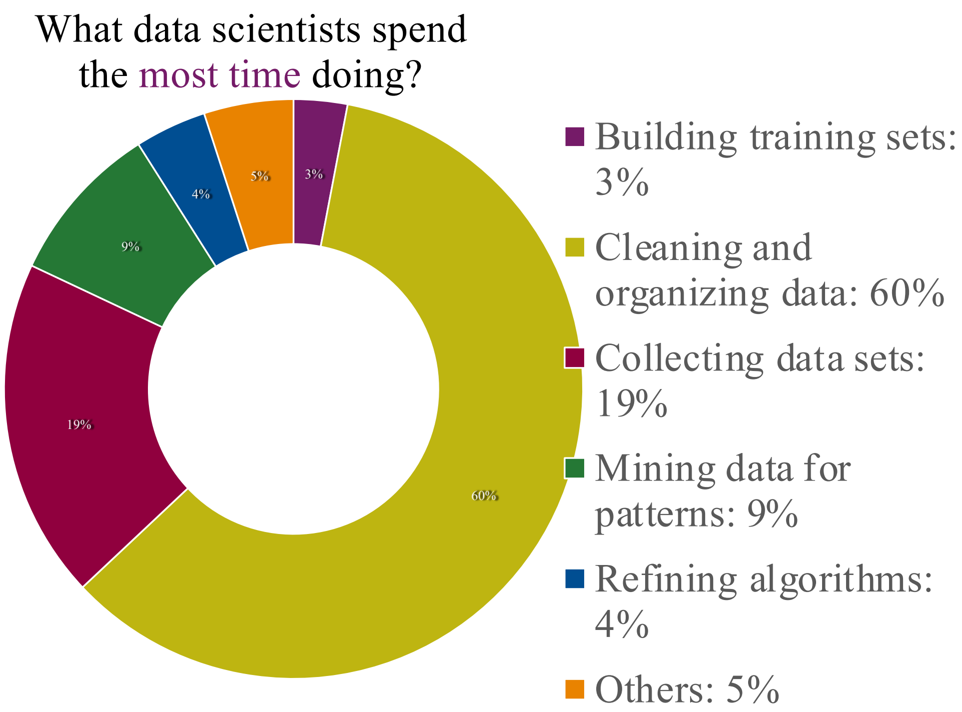


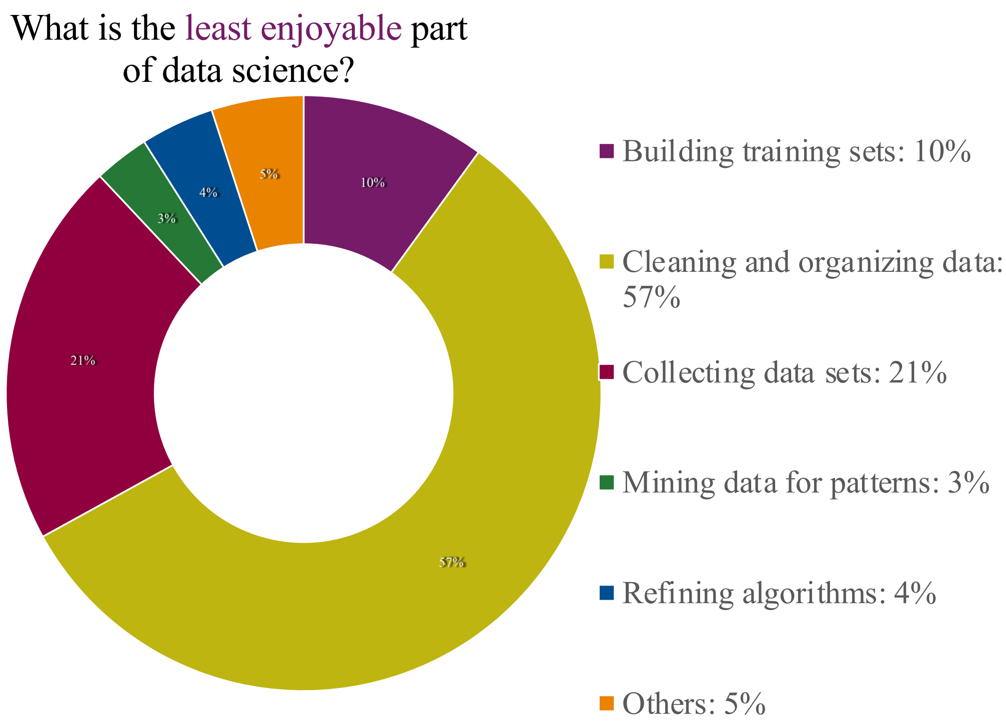
Data Science Pipeline



UNIVERSITY OF AMSTERDAM
Faculty of Science

Data preparation in reality



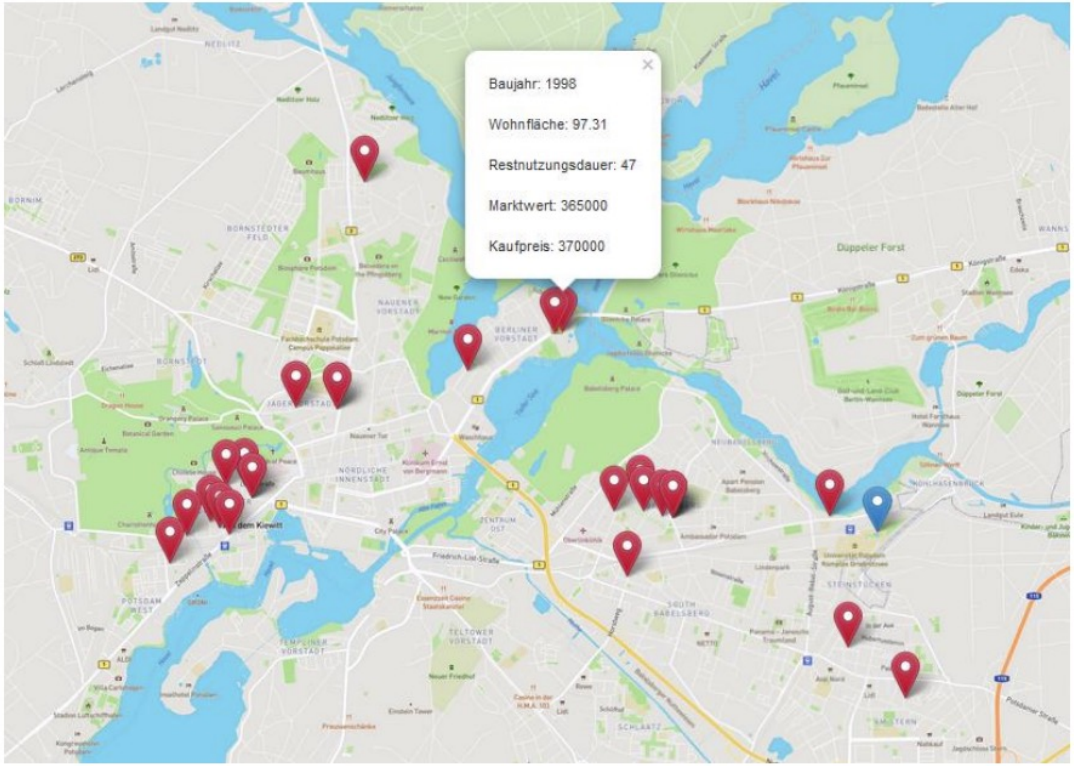


“Cleaning Data: Most Time-Consuming, Least Enjoyable Data Science Task”, Gil Press, Forbes, March 23rd, 2016
http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/

1
3

Real World meets Data Quality

Residential real estate valuation



Vergleichswert

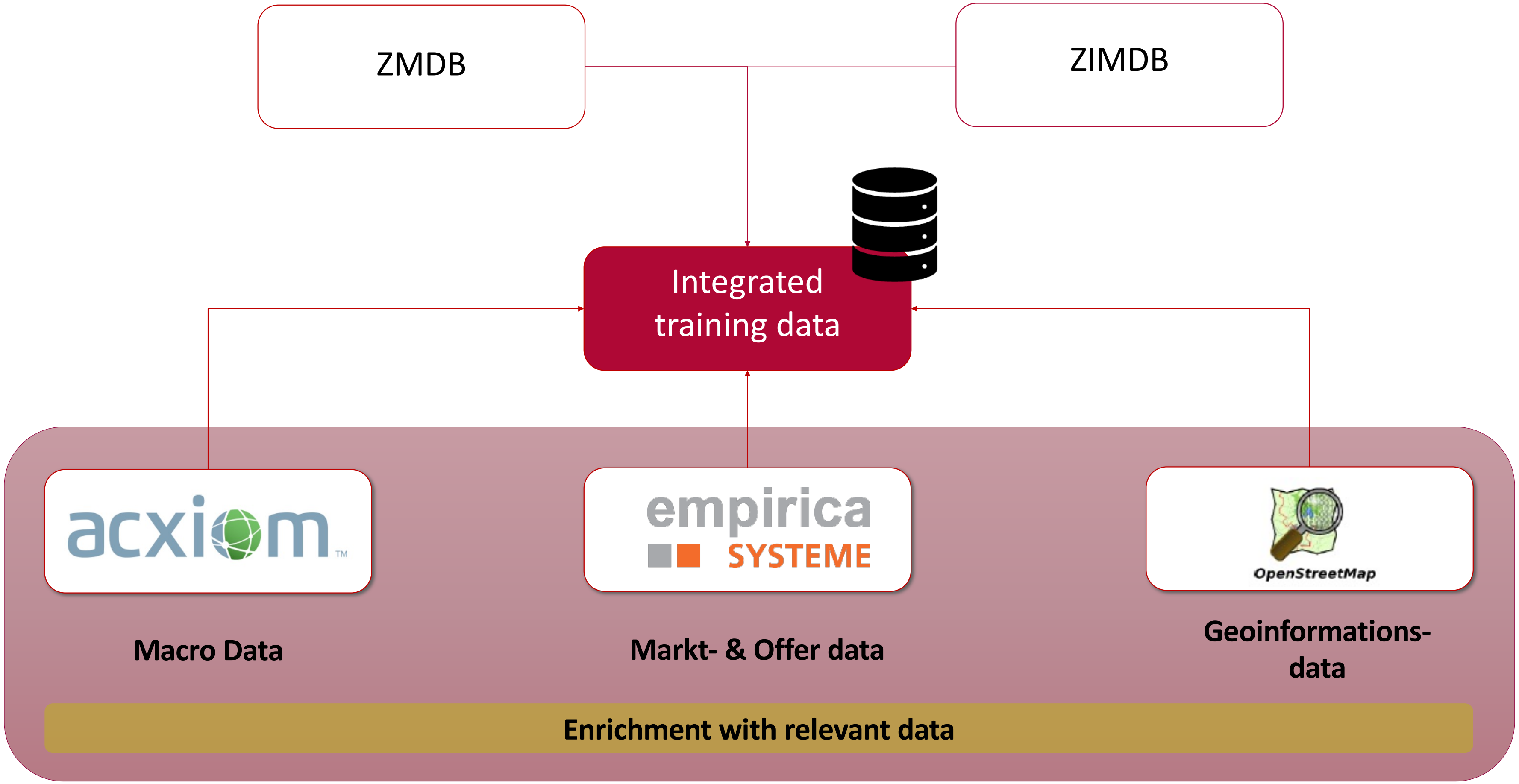
Marktwert	163560 €
Sicherheitsabschlag	20 %
Beleihungswert	**130690 €**
Spannbreite	147204 € bis 196271 €

Es wurden genug Vergleichsobjekte gefunden.

Es gibt genug Objekte zur Berechnung einer Spannbreite.

University of Amsterdam
Faculty of Science

Real World meets Data Quality



Real World meets Data Quality

- Technical problems
  - Json dump from a relational database
    - NonSQL as base for analysis!
  - Sparkasse banks are non-central
    - Different schemata

- Relevance (syntactic and semantic)
  - object and field wise.

- Duplication:
  - same houses
  - same attributes across collections in ZIMDB

UNIVERSITY OF AMSTERDAM
Faculty of Science

Real World meets Data Quality

Missing values

Quantity	%	Coverage in %
#4	5,33%	≤ 10%
#3	4,00%	10,01% - 30%
#9	12,00%	30,01% - 50%
#19	25,33%	50,01% - 90%
#39	52,00%	90,01% - 100%

The coverage of the relevent fields in ZIMDB

```
if gutachtenart = "Kurzgutachten"
        integrated_marktwert :=
kurzgutachten.ergebnisMarktwertGerundet
else if gutachtenart = "Vollgutachten"
        if ableitungsGrundlageMwt = "Sachwert"
                integrated_marktwert := ergebnisSachwertMwt
        else if ableitungsGrundlageMwt= "Ertragswert"
                integrated_marktwert := ertragswertGerundetMwt
        else if ableitungsGrundlageMwt= "Vergleichswert"
                integrated_marktwert := marktwert
        else
                integrated_marktwert := „“
```

Attempt to impute our target variable for learning

UNIVERSITY OF AMSTERDAM
Faculty of Science

Real World meets Data Quality

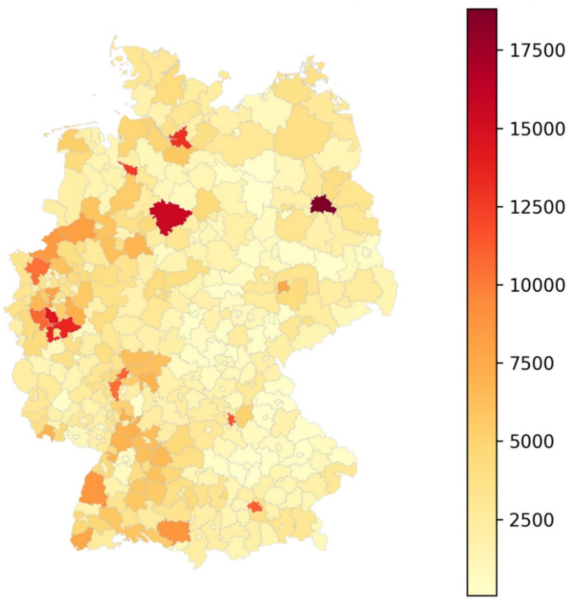
- Inconsistent representation
    - Manual Mappings

Standard	Actual values
Einfamilienhaus	Ein- / Zweifamilienhäuser
	Ein-/Zweifamilienhaus
	Ein-/Zweifamilienwohnhaus
Zweifamilienhaus	Zweifamilienwohnhaus
Eigentumswohnung	Eigentumswohnung(en)
	Wohnungsbau
Wohngrundstück	Wohnobjekt Eigennutzung
	Teileigentum Wohnen
	Wohnimmobilie Bremen
	Wohnungseigentum
	Wohnimmobilie
	Wohn- und Geschäftsimmobilie
	Wohneigentum

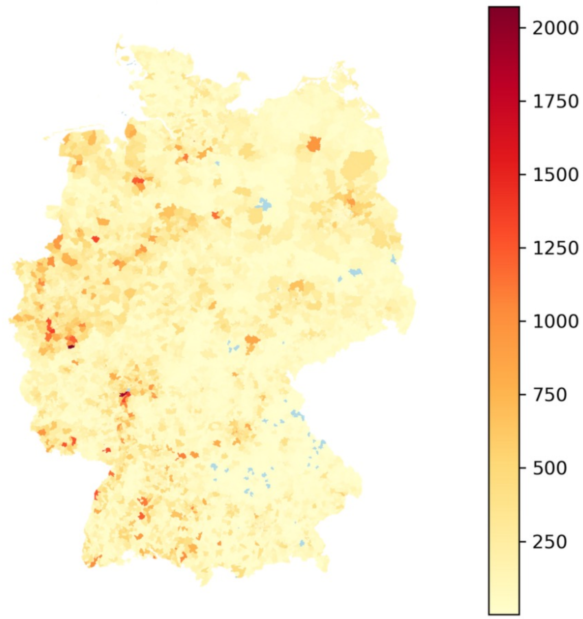
Standard	gut	mittel	schlecht
Vorgefundene Werte	1 - exzellent	Weniger gut	einfach
	bevorzugt	mäßig (4)	sehr schlecht (6)
	2 - sehr gut	mäßig	ungünstig
	Sehr gut	befriedigend	katastrophal
	gut (2)	mittel-deaktiv	leicht unterdurchschnittlich
	hervorragend	ausreichend	starke Beeinträchtigung
	überdurchschnittlich	mittel	schlecht
	4 - überdurchschnittlich	durchschnittlich	8 - schlecht
	gut	7 - mäßig	unterdurchschnittlich
	sehr schlecht	5 - durchschnittlich	weniger gut
	Gut	normal	schlecht (5)
	exzellent	befriedigend (3)	9 - sehr schlecht
	Bevorzugt		6 - unterdurchschnittlich
	sehr gut (1)		---einfach---
	gehoben		weniger Gut
	Bevorzugt - sehr gut		Schlecht
	leicht überdurchschnittlich		10 - katastrophal
	3 - gut		ungenügend
	sehr gut		sehr einfach
	beste		
	Hervorragend		
	gut		

Real World meets Data Quality

Representivity



per county (Kreis) (absolute)

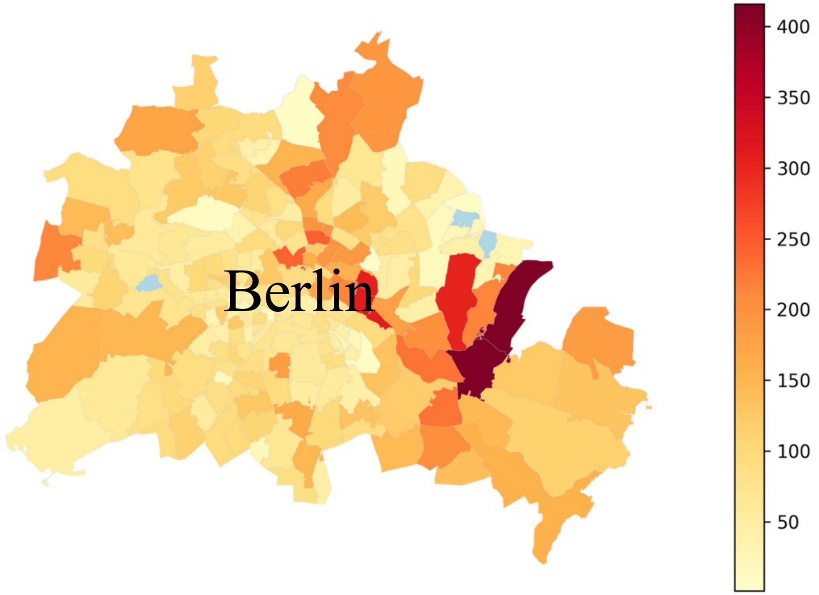


per zip code (absolute)

Blue zip code: No data points

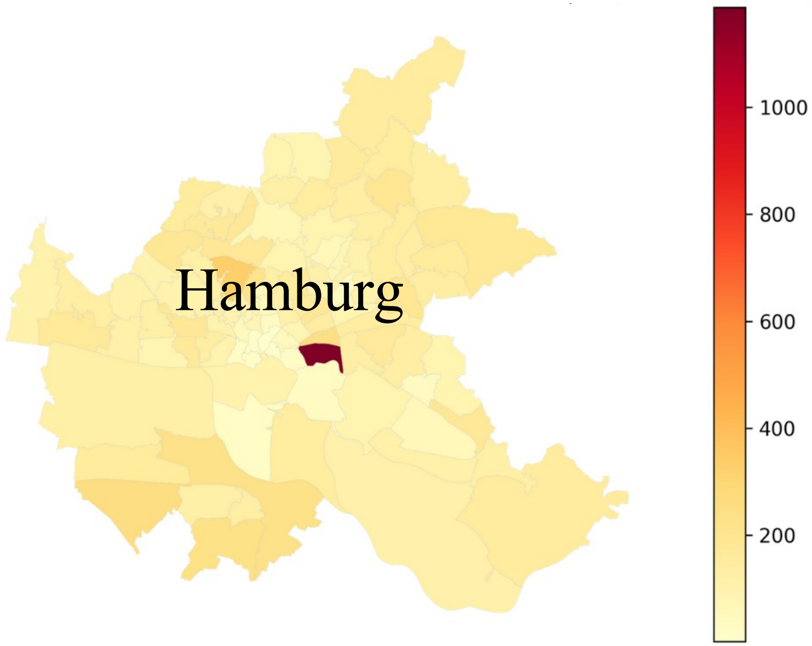
Real World meets Data Quality

Outliers



per county (Kreis) (absolute)

Blue zip code: No data points

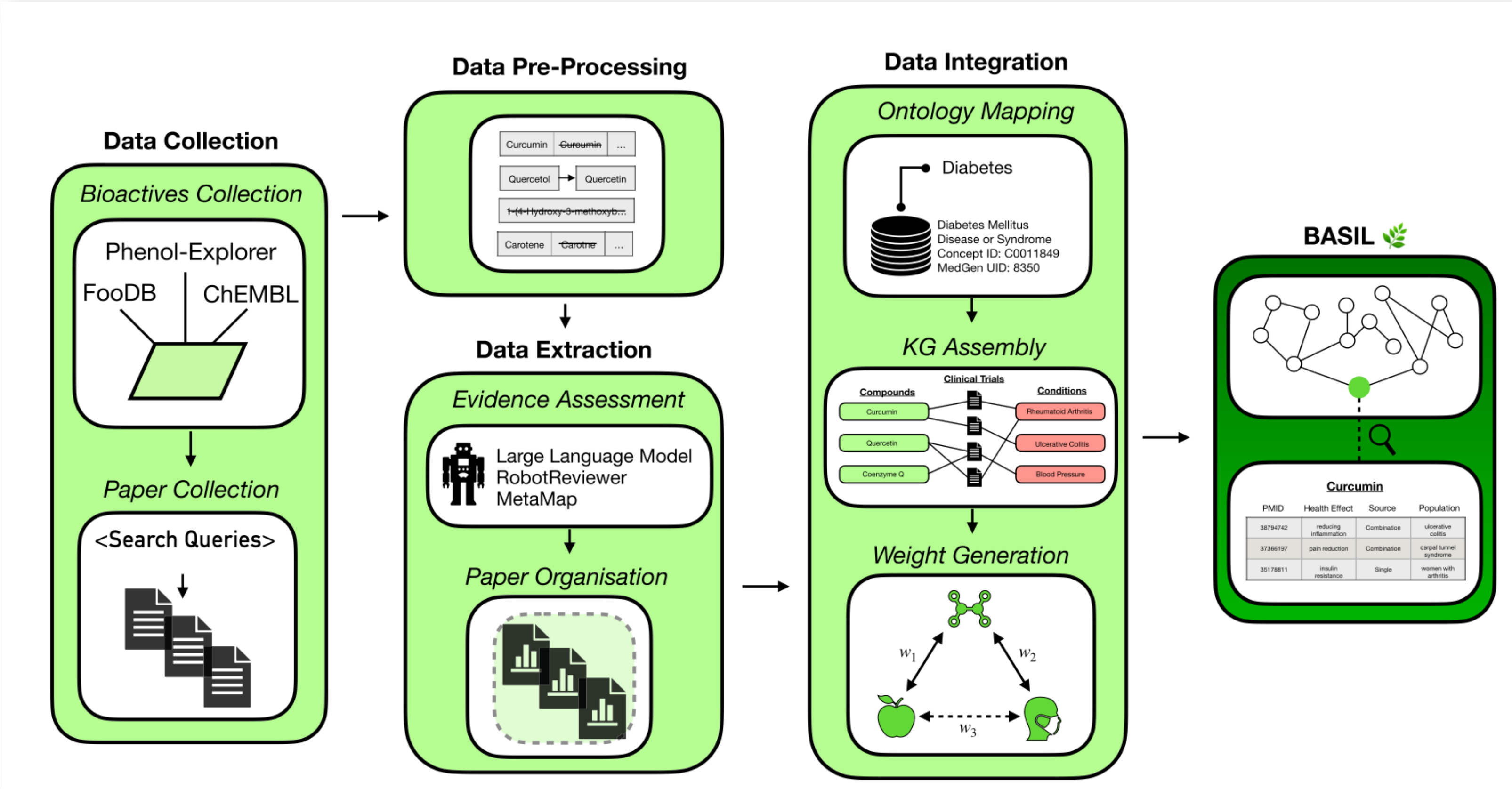


per zip code (absolute)

High number in the middle: Headquarters in Hamburg

UNIVERSITY OF AMSTERDAM
Faculty of Science

Real World meets Data Quality – Medical data



BASIL DB: BioActive Semantic Integration and Linking Database: David Jackson, Paul Groth, and Hazar Harmouch. Journal of Biomedical Semantics, 2025: https://rdcu.be/eAFql

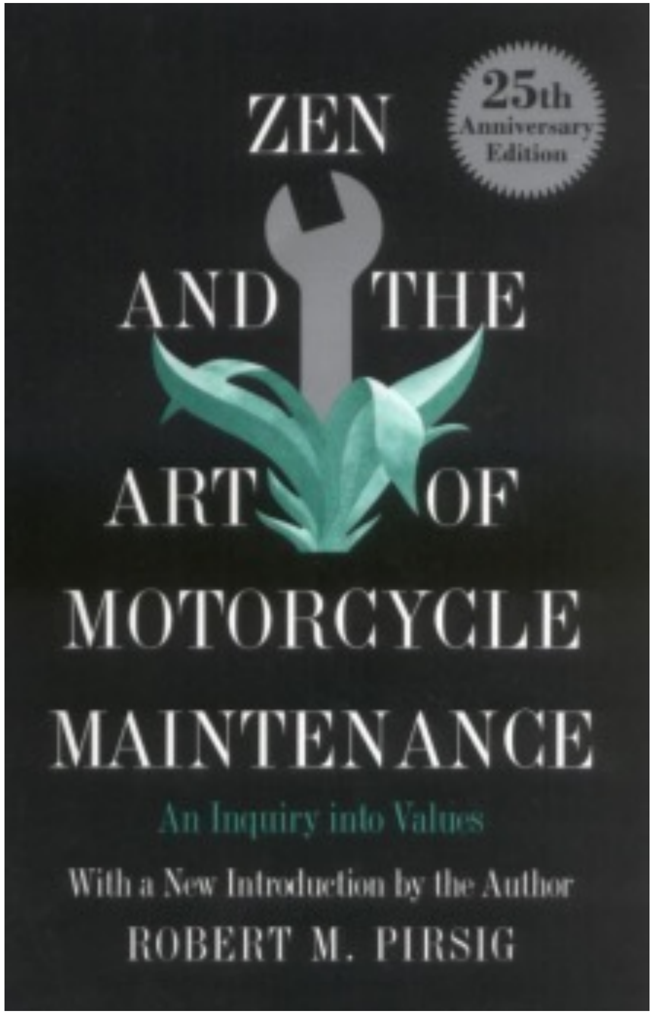
UNIVERSITY OF AMSTERDAM
Faculty of Science

Agenda

1. **Data and Information Quality Research**
2. Data Quality and AI Systems
3. Cleaning For ML
4. Data Quality Assessment and ongoing projects

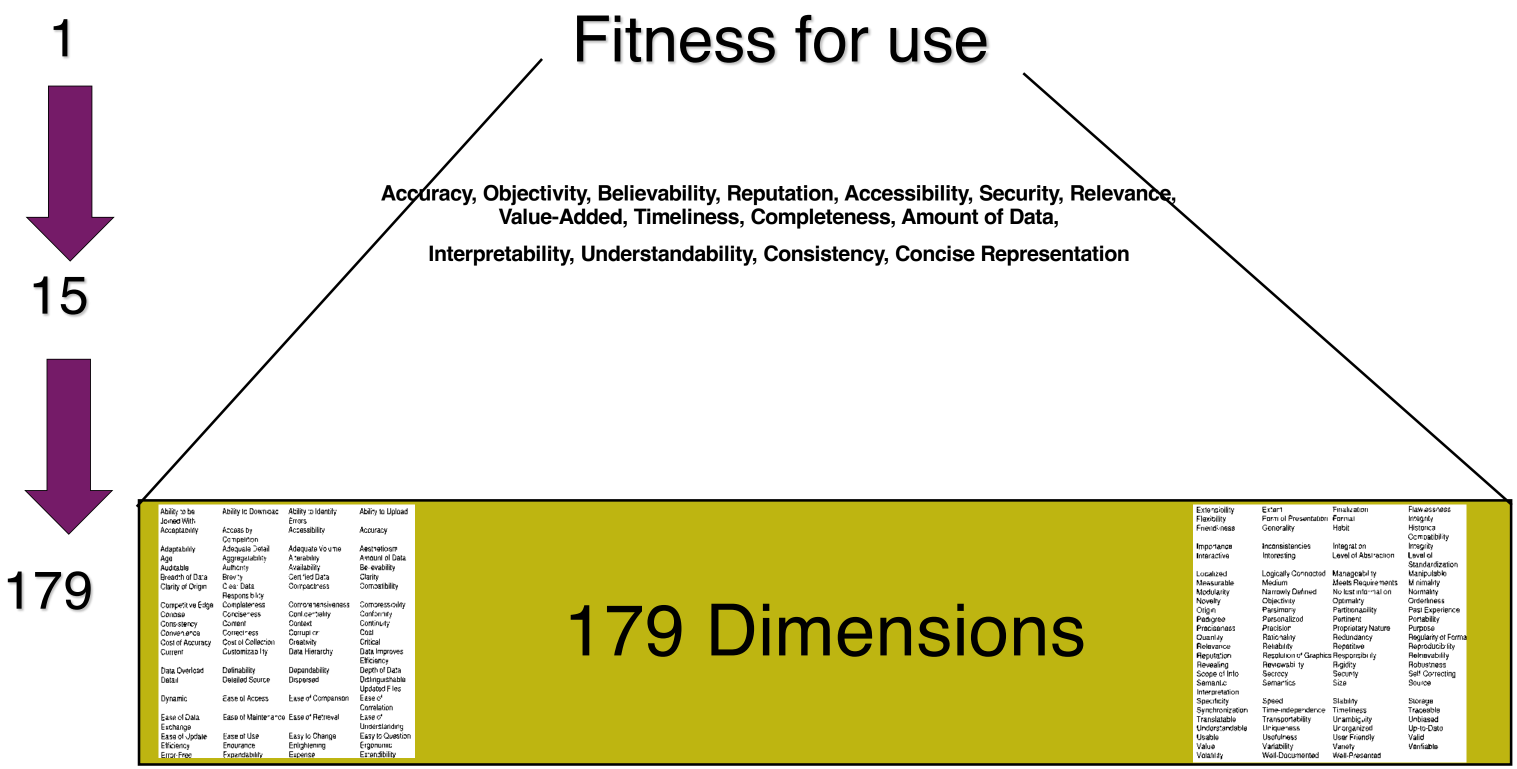
Quality

"Even though quality cannot be defined, you know what it is."
Robert Pirsig



UNIVERSITY OF AMSTERDAM
Faculty of Science

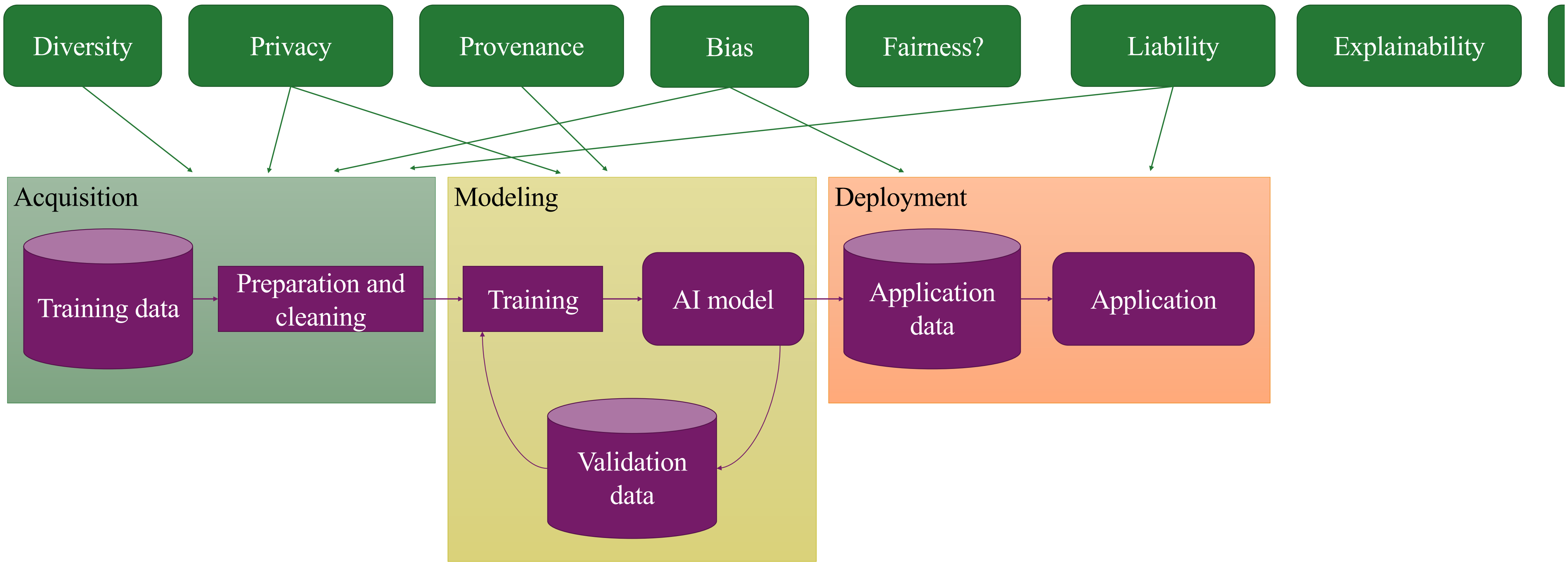
Zooming into Information Quality



Wang, R. Y. & Strong, D. M.
Beyond Accuracy: What data quality means to data consumers
Management of Information Systems, 1996*, 12(4)*, 5-34

University of Amsterdam
Faculty of Science

New AI-specific Data Quality Dimensions



UNIVERSITY OF AMSTERDAM
Faculty of Science

A Data Quality Glossary, Sedir Mohammed et. al, Jan 2024
https://doi.org/10.5281/zenodo.10474880

28 DQ Dimensions

Uniqueness
Relevance
Privacy
Understandability
Accuracy
Consistent representation
Security
Believability
Added value
Fairness
Precision
Diversity
Representativity
Reputation
Documentation
Transparency
Consistency
Balance
Reliability
Completeness
Traceability
Portability
Accessibility
Amount
Recoverability
Timeliness
Compliance
Efficiency
Ease of manipulation

UNIVERSITY OF AMSTERDAM
Faculty of Science

Agenda

1. Data and Information Quality Research
2. **Data Quality and AI Systems**
3. Cleaning for ML
4. Data Quality Assessment

Empirical Measurement of the Effects of Poor Data Quality on ML Results

Pollutions

- Consistent representation
- Completeness
- Feature accuracy
- Target accuracy
- Uniqueness
- Target balance

Scenarios

- Pollute only training data
- Pollute only test data
- Pollute training and test data

Runs

- 5 runs, average

Tasks and algorithms

- Classification
  - LogR, SVM, DT, GB, KNN, MLP
- Clustering
  - GM, k-Means, k-Prototypes, AC, OPTICS
- Regression
  - LR, RR, DT, RF, GB, MLP, TabNet

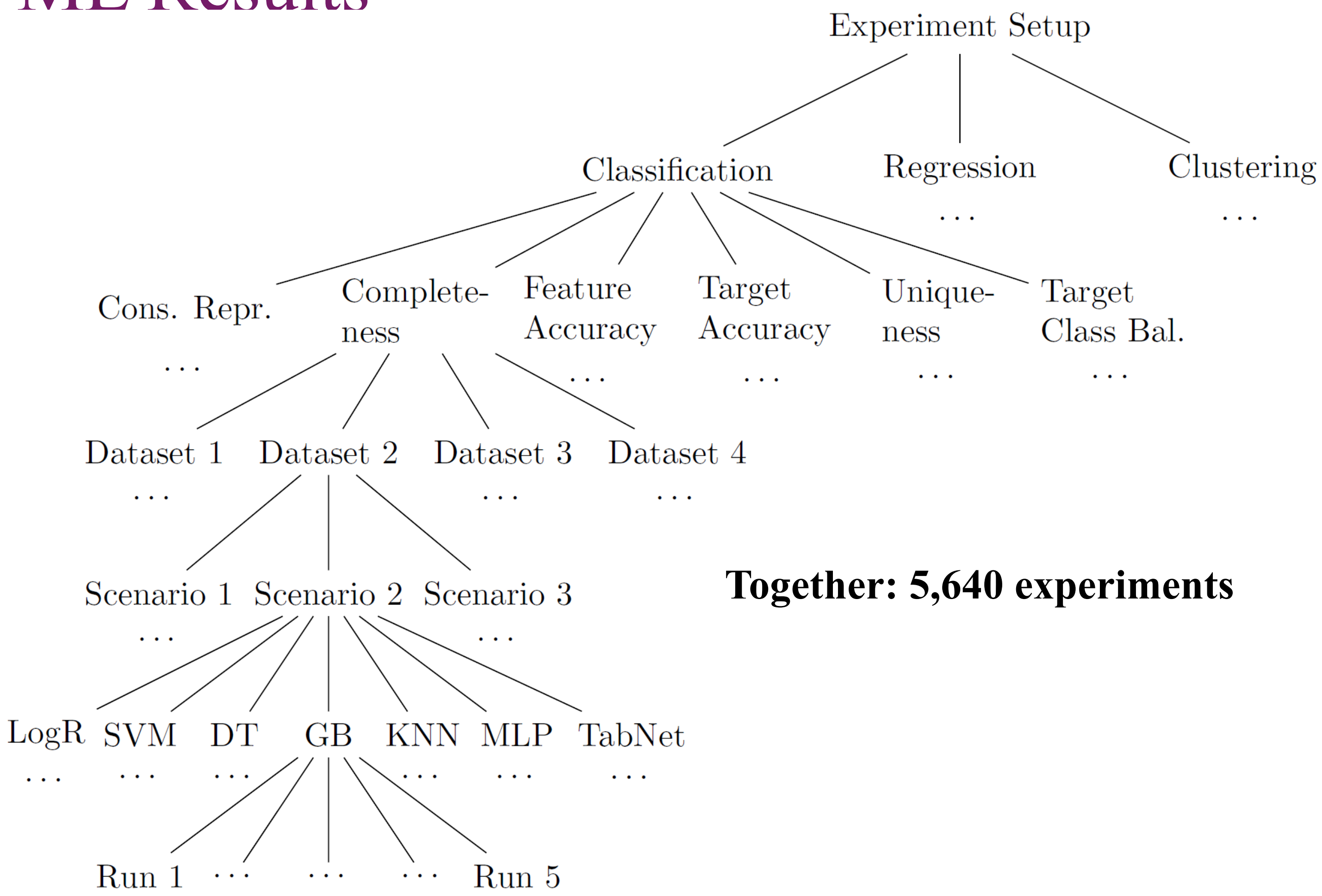
Datasets

- TelcoChurn, GermanCredit, Contraceptive, COVID
- Houses, IMDB, Cars
- Bank, Covertype, Letter

The Effects of Data Quality on Machine Learning Performance, Sedir Mohammed et. Al. , Information Systems (2025)

UNIVERSITY OF AMSTERDAM
Faculty of Science

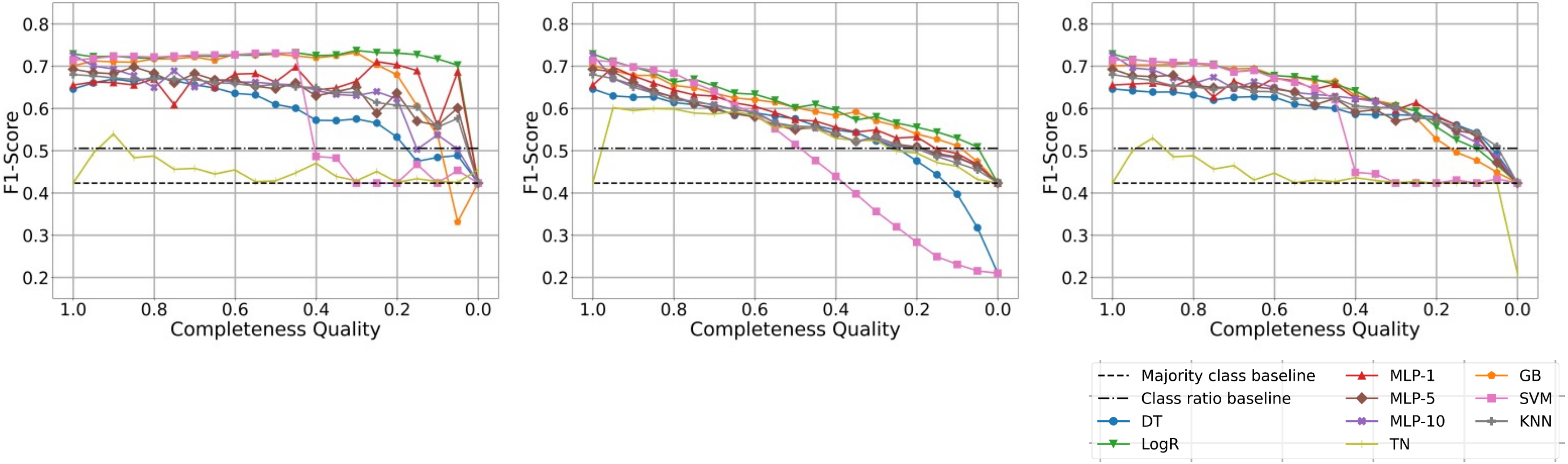
Empirical Measurement of the Effects of Poor Data Quality on ML Results



Together: 5,640 experiments

UNIVERSITY OF AMSTERDAM
Faculty of Science

Example Results



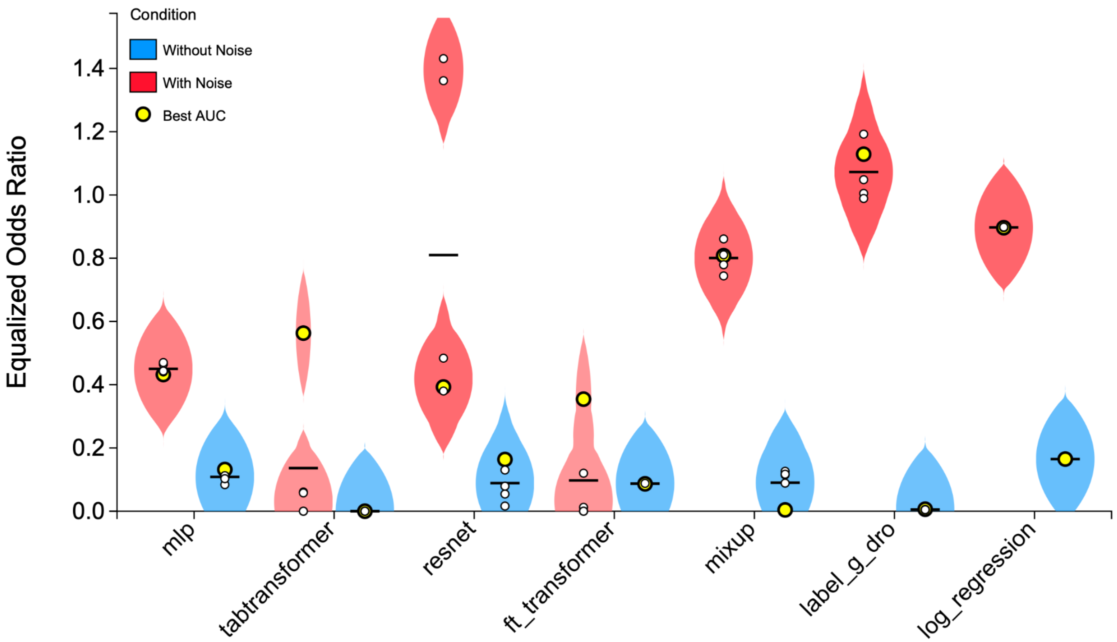
Average F1-Score for Classification of the Telco-Churn dataset

Qualitative Trends

The effect of data quality dimensions per ML task. ✓: low effect, ○: moderate effect, ×: high effect.

	Consistency	Completeness	Feat.-Accuracy	Tar.-Accuracy	Uniqueness	Class Balance
Classification	✓	×	×	×	✓	○
Regression	✓	×	×	×	✓	○
Clustering	✓	×	×	✓	✓	✓

The Impact of Labels Quality on ML Robustness and Fairness

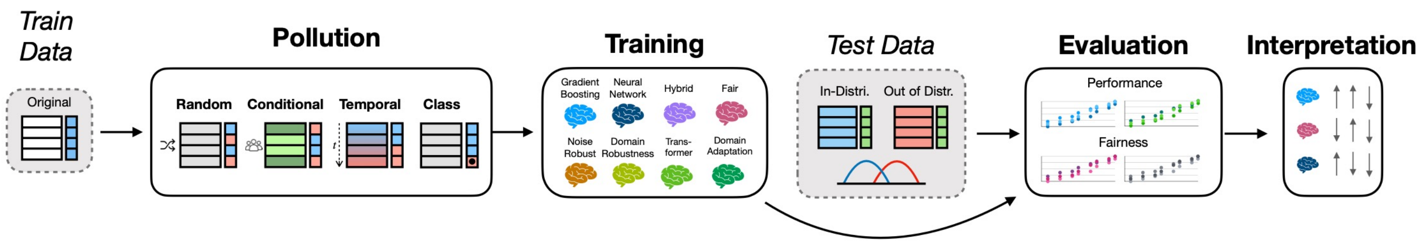


Fault Lines: Benchmarking the Impact of Label Data Quality on ML Robustness and Fairness [Experiment, Analysis & Benchmark] Under revision, PVLDB

UNIVERSITY OF AMSTERDAM
Faculty of Science

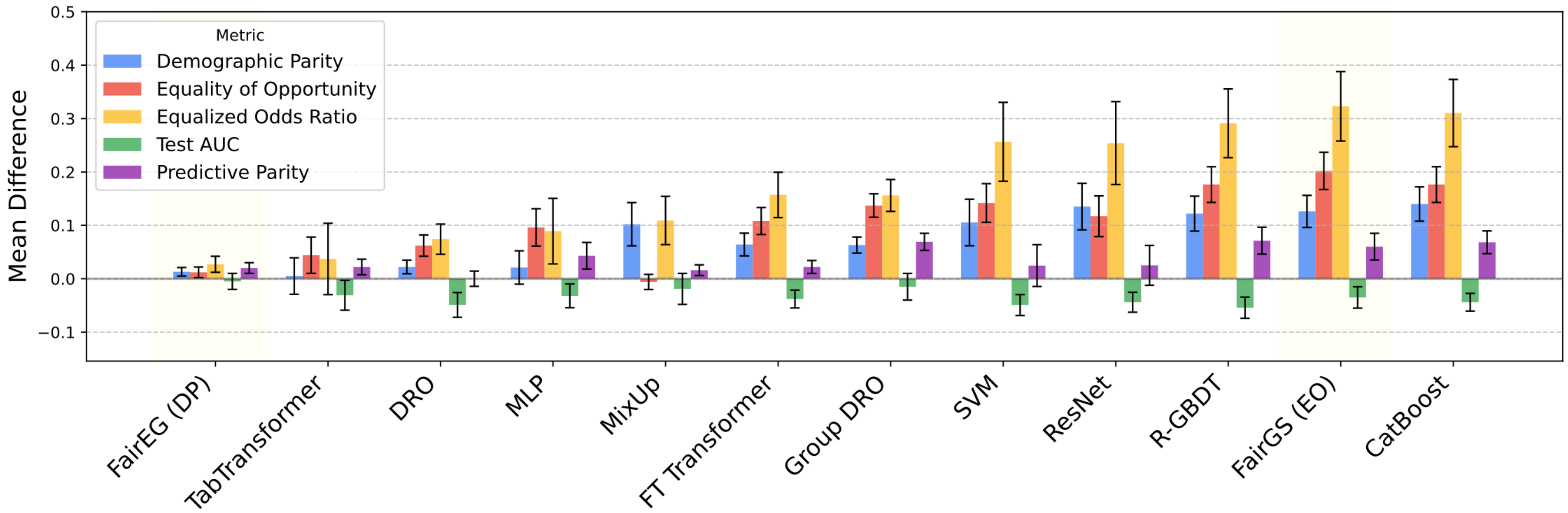
Introducing Fault Lines Benchmark

- **15 diverse tabular datasets** spanning healthcare, finance, and social outcome
- **22 state-of-the-art models** including boosting, transformers, and fairness-aware approaches
- Novel noise types that mirror reality
    - **Random**: Traditional uniform label flipping
    - **Biased**: Feature-dependent + class-conditional noise targeting specific subgroups
    - **Correlated / Concatenated**: Multiple feature interactions
    - **Temporal**: Time-dependent corruption patterns



UNIVERSITY OF AMSTER
Faculty of Science

Results



- Striking Asymmetry in Robustness
  - **<10% biased noise** causes substantial performance degradation
  - Up to **700% increase** in fairness disparities
  - Performance may appear stable, yet fairness still degrades

- Model selection, noise type, and dataset characteristics (size, imbalance ratio, subgroup sizes) tightly interlinked

Usecases

- Data cleaning pipeline evaluation - Test your methods against realistic noise
- Model selection guidance - Choose architectures based on expected bias patterns
- Fairness monitoring - Detect when noise undermines equity

The Bottom Line

Even small amounts of systematic bias in labels can undermine both robustness and fairness: requiring targeted interventions beyond traditional data cleaning.

UNIVERSITY OF AMSTERDAM
Faculty of Science

Agenda

1. Data and Information Quality Research
2. Data Quality and AI Systems
3. **Cleaning for ML**
4. Data Quality Assessment

COMET: Cleaning Optimization and Model Enhancement Toolkit

- COMET guides the user stepwise through the cleaning process.
  - Considering a benefit-cost ratio, which feature should be cleaned next?

Higher prediction accuracy increase

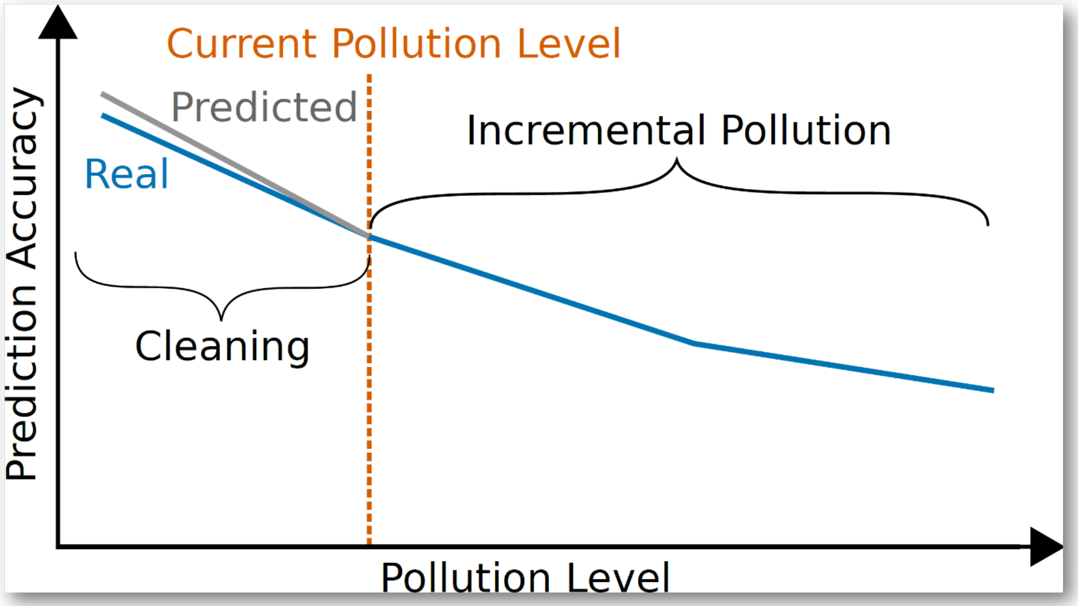
Lower cleaning costs

- Error type-agnostic and ML algorithm-agnostic

Step-by-Step Data Cleaning Recommendations to Improve ML Prediction Accuracy, Sedir Mohammed , Felix Naumann , and Hazar Harmouch, *EDBT 2025*, Feb 2025

COMET

Extrapolating cleaning trends out of dirty data

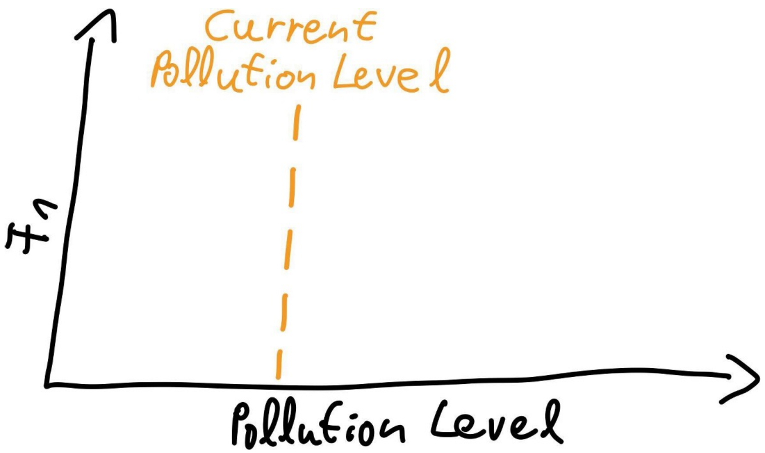


UNIVERSITY OF AMSTERDAM
Faculty of Science

COMET

Error Types

Input Data

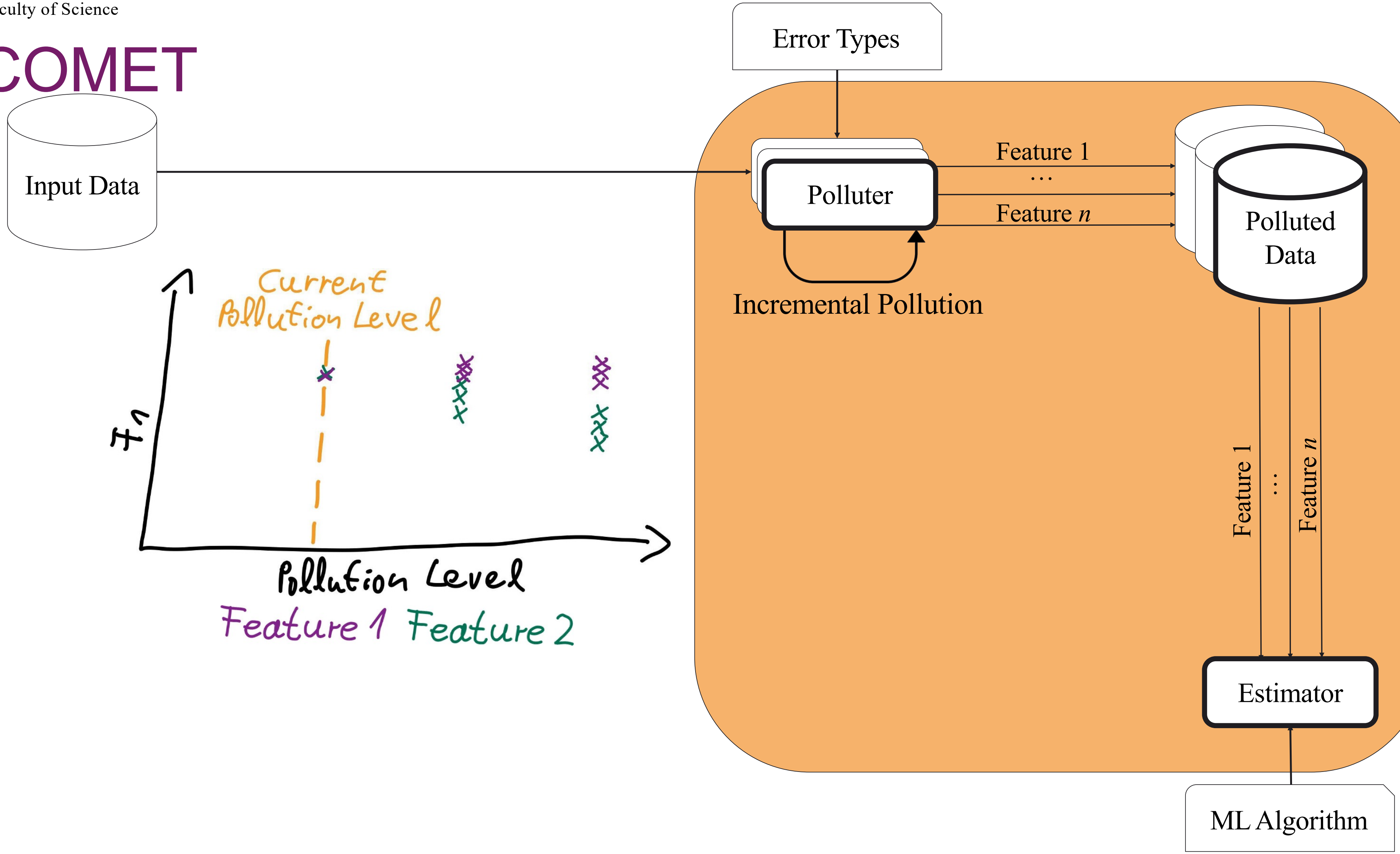


ML Algorithm

4
2

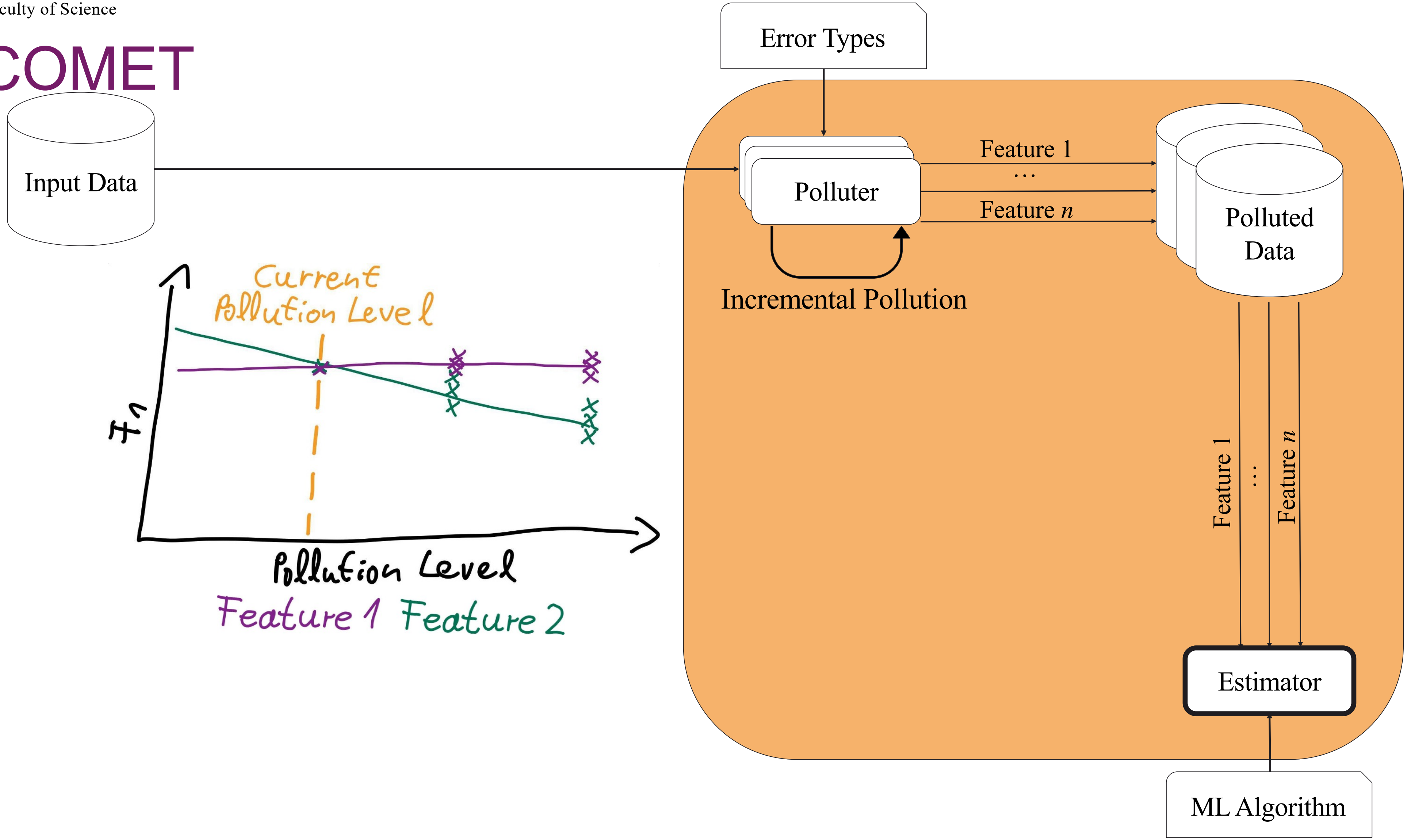
University of Amsterdam
Faculty of Science

COMET



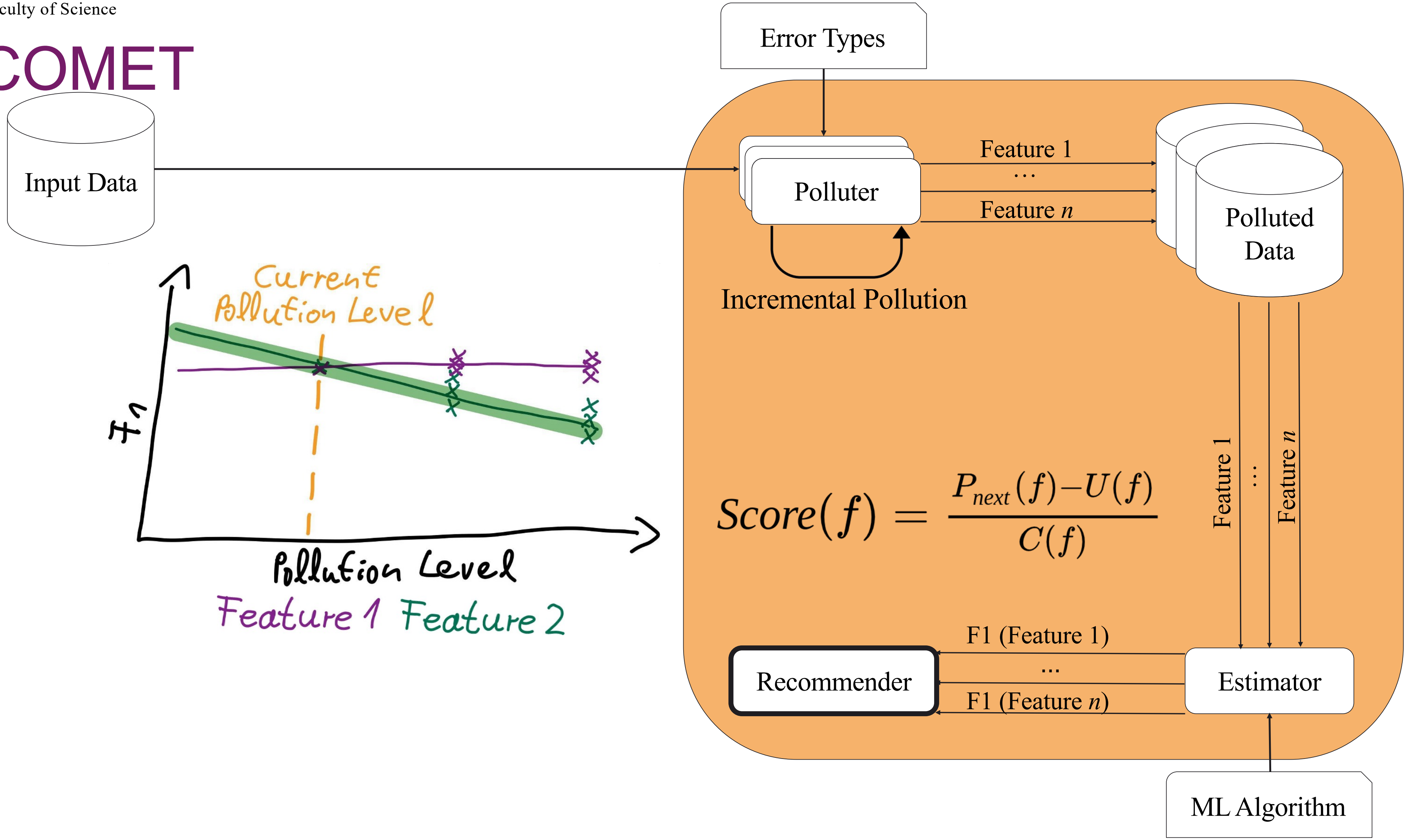
4
3

COMET


Input Data
Error Types
Polluter
Feature 1
...
Feature n
Polluted Data
Incremental Pollution
Feature 1
...
Feature n
Estimator
ML Algorithm
Current Pollution Level
F1
Pollution Level
Feature 1
Feature 2

UNIVERSITY OF AMSTERDAM
Faculty of Science

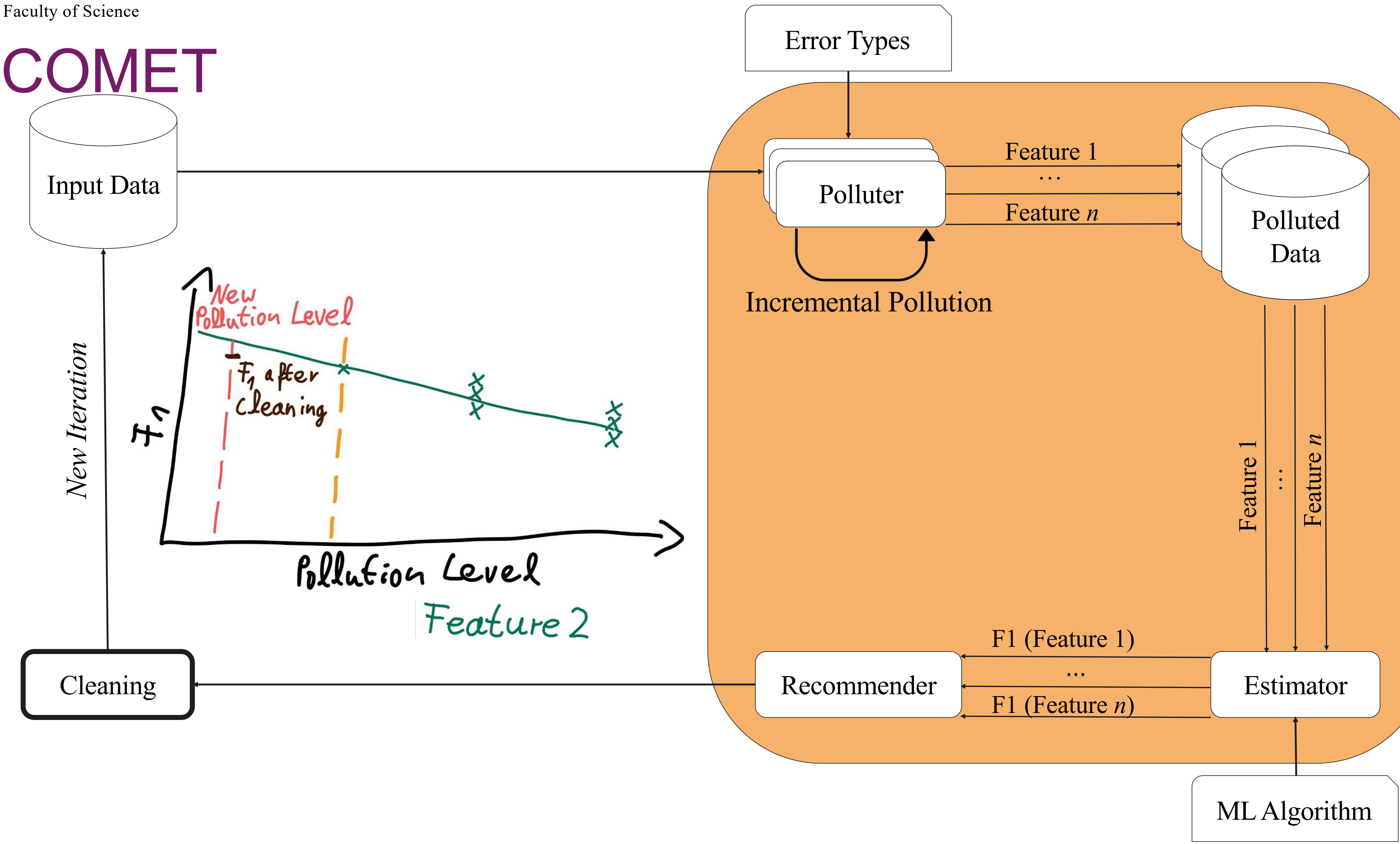
COMET



4
5

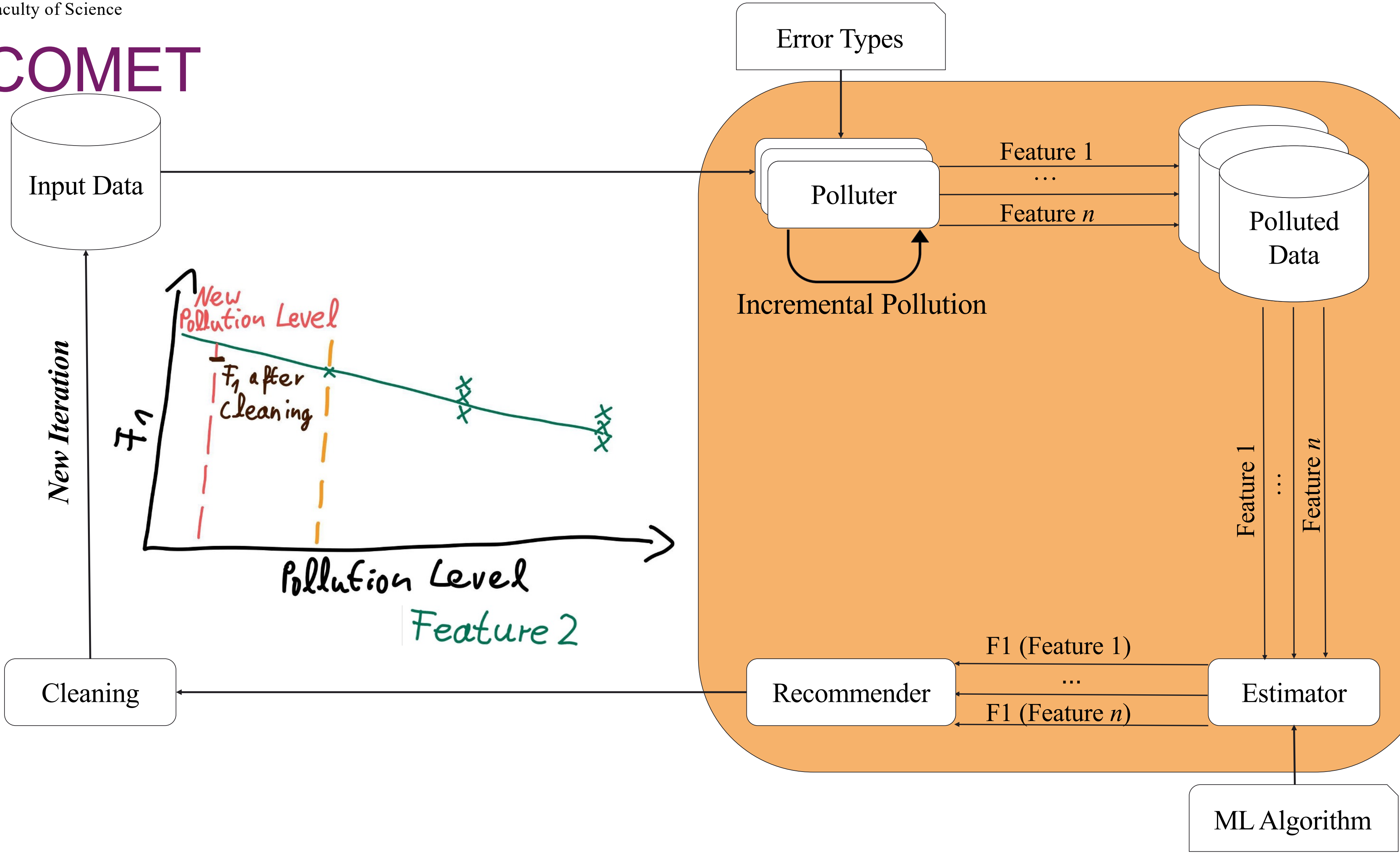
University of Amsterdam
Faculty of Science

COMET



University of Amsterdam
Faculty of Science

COMET

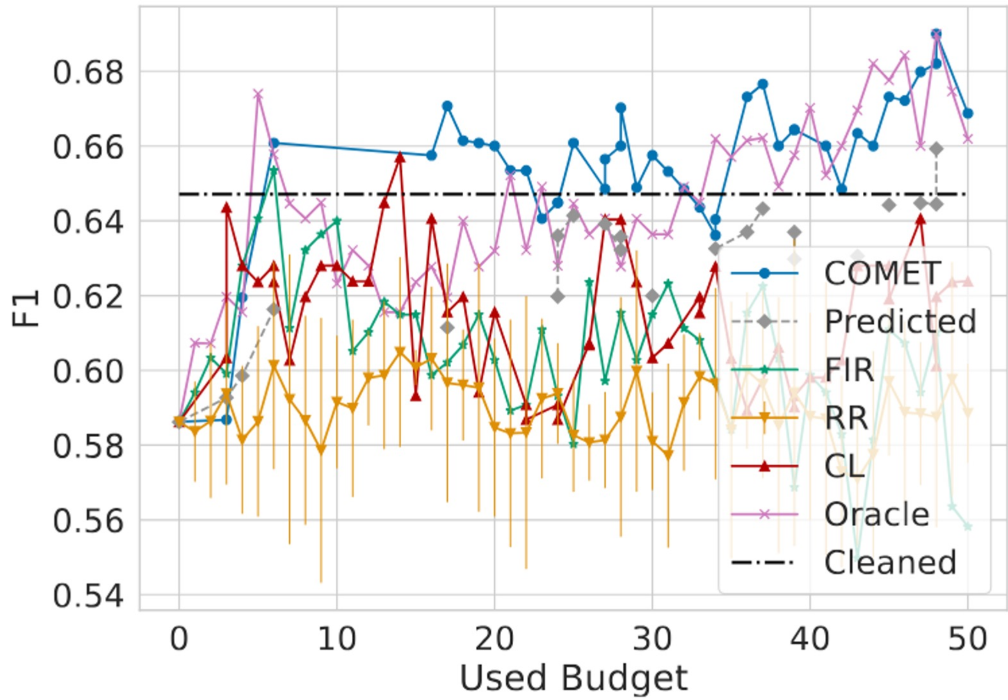


Experiment Setting

- Baselines
  - Random recommendations (RR)
  - Feature importance-based recommendations (FIR)
  - Light COMET (CL)
  - ActiveClean (AC)
  - Oracle
- We tested COMET with:
  - Error types: Missing values, Gaussian noise, categorical shift, scaling
  - ML Algorithms: Support Vector Machine (SVM), *K*-Nearest Neighbour (KNN), Multi-Layer Perceptron (MLP), Gradient Boosting (GB); AC: Linear Regression (LIR), Logistic Regression (LOR), AC-SVM
- Datasets 7 datasets (3 with given ground truth)

University of Amsterdam
Faculty of Science

Performance comparison for a single Error Type

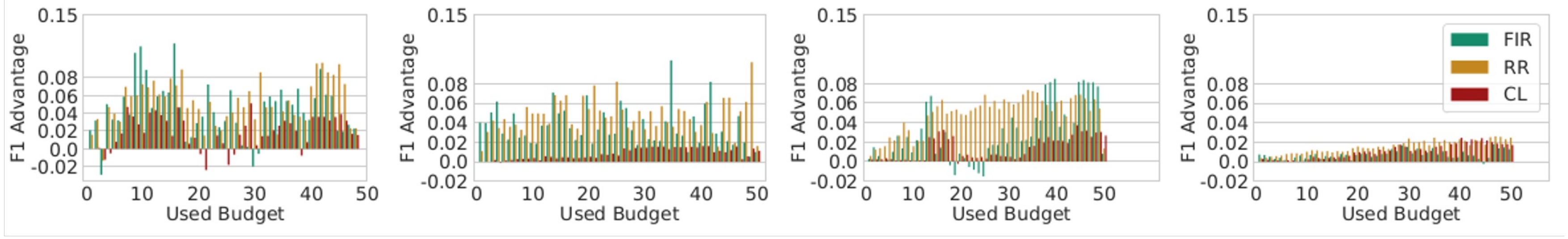


SVM - Categorical shift error

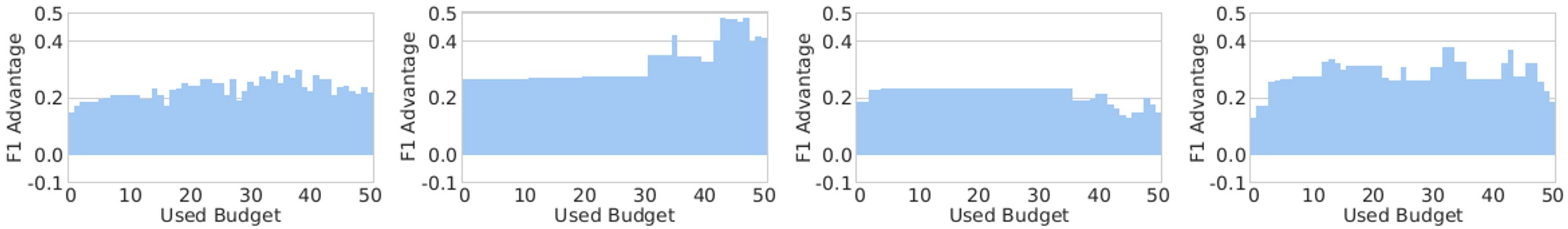
RR - Random recommendations; FIR - Feature importance-based recommendations; CL - Light COMET

Performance for multiple Error Types

Comparison to FIR, RR, CL



Comparison to AC



CMC Churn EEG S-Credit

UNIVERSITY OF AMSTERDAM
Faculty of Science

Holistic Data Cleaning for Machine Learning (Ongoing)

Table 4. Comparison of core repair methods

Aspect	HoloClean [59]	Com
ML Perf. Opt.	✗	✓
External Data	✓	✗
Cost Awareness	✗	✓(m
Recommendations	✗	✓
Automation / Human Role	Full / None	Semi
Validation Set	✗	✓
Theory Base	DB constraints	Info
Multiple Signals	✓	✗
Real-time Adaptation	✗	✓
Budget Constraints	✗	✓
Early Termination	✓	✓
Uncertainty Quantification	✓	✓
Domain Reqs.	IC	Non

Table 5. Comparison of core LLM-based holistic data cleaning approaches.

Aspect	GIDCL [73]	IterClean [49]	LLMClean [10]	UniDM [54]	AutoDCWorkflow† [35]	LLMAgents†
LLM Role	Creator-critic + rule gen	Detector, verifier, repairer	OFD extraction & prompting	Retrieval, parsing, prompting	Workflow planner	Interactive agent
Approach	GNN + LLM + PLM	Iterative multi-role LLM	OFD rules from LLM	Unified formalization	Purpose-driven workflow	Iterative explore–clean–evaluate
Manual Input	20 tuples	5 tuples	None	Task parameters	Purpose only	Prompt + target
Interpretability	Interpretable rules	Limited	Interpretable OFDs	Limited	Explicit workflows	Partial (code visible)
Error Coverage	Syntactic, semantic	Outliers, violations, patterns	Dependency & missing	Task-dependent	Duplicates, missing, format	Num. shift, NaN, categorical shift
Handles Dependencies	✓(graph)	Limited	✓(OFD)	✗	Limited	Limited
Multi-Table	Limited	✗	Limited	✓	✗	✗
ML Optimization	✗	✗	✗	✗	✗	✓
Iterative Process	✓	✓	✗	✗	✓	✓
Verification	Critic role	Verifier role	None	None	Data quality report	ML model score

† Methods not yet peer-reviewed at the time of assessment.

Agenda

1. Data and Information Quality Research
2. Data Quality and AI Systems
3. Cleaning for ML
4. **Data Quality Assessment**

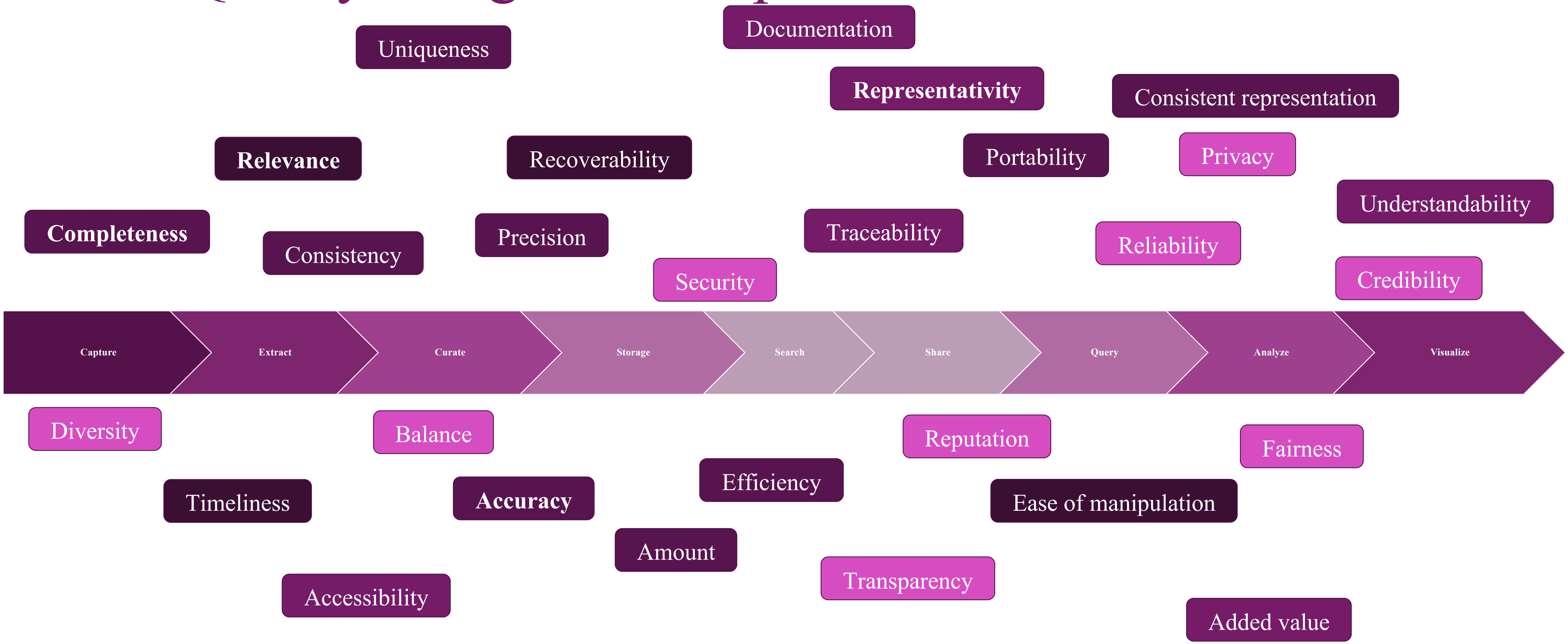
Assessing Data Quality

Recoverability
Portability
Privacy
Uniqueness
Accuracy
Consistent representation
Security
Documentation
Added value
Believability
Fairness
Precision
Balance
Transparency

Measure
Check
Assess

Timeliness
Amount
Consistency
Reliability
Accessibility
Relevance
Traceability
Completeness
Ease of manipulation
Efficiency
reputation
Compliance
Diversity
Understandability
Representativity

Data Quality along the AI Pipeline



University of Amsterdam
Faculty of Science

European Parliament

European AI Act Article 10 (3): Data and Data Governance

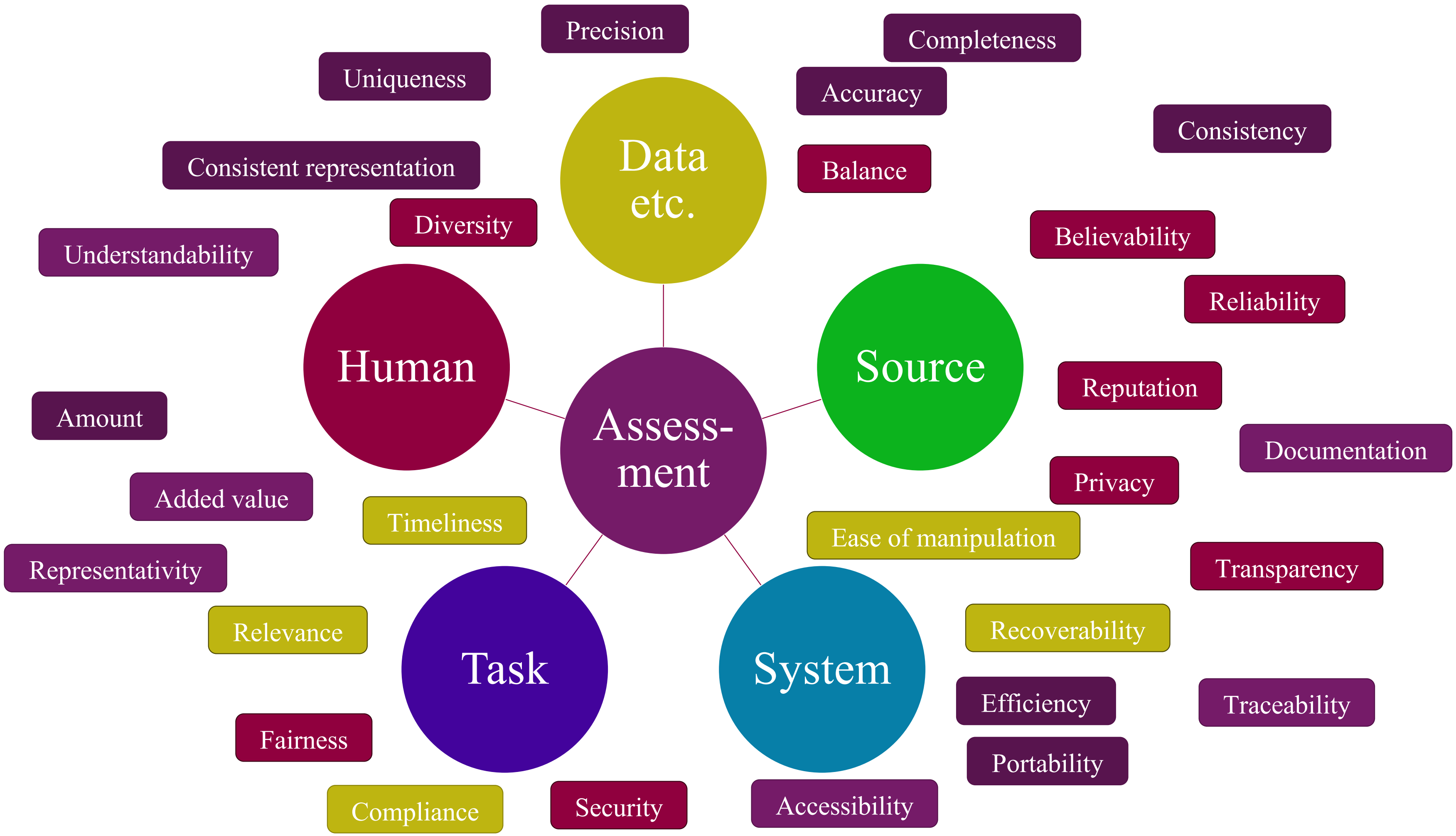
- High-quality data and access to high-quality data plays a vital role in providing structure and in ensuring the performance of many AI systems, especially when techniques involving the training of models are used, with a view to ensure that the high-risk AI system performs as intended and safely and it does not become a source of discrimination prohibited by Union law.
- High-quality data sets for training, validation and testing require the implementation of appropriate data governance and management practices.
- **Data sets for training, validation and testing, including the labels, should be relevant, sufficiently representative, and to the best extent possible free of errors and complete in view of the intended purpose of the system.**
- The data sets should also have the appropriate statistical properties, including as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used, with specific attention to the mitigation of possible biases in the data sets […].

https://artificialintelligenceact.eu/article/10/

University of Amsterdam
Faculty of Science

The Five Facets of Data Quality Assessment. Sedir Mohammed , Lisa Ehrlinger , Hazar Harmouch, Felix Naumann , and Divesh Srivastava, SIGMOD Records, Apr 2025

Ingredients for DQ Assessment: Five Facets



Assessment Examples

Completeness

- Values vs. rows vs. columns
- Nulls vs. disguised missing values
- External data needed
- Semantically challenging

Representativity

- vs. balance vs. diversity
- Presence of every value combination
  - Existing values vs. all values
- Computationally challenging

Free-of-errors / Correctness

- Error detection
  - Count at value or row-level
- Business rules
  - Patterns, dependencies, data-types
- Outlier detection
- Validation with external data

Relevance

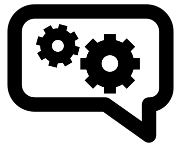
- …

Understandability

- …

Further Challenges for DQ Assessment

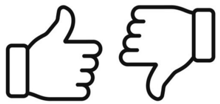
- Ambiguity
  - Many attempts to compile and define DQ dimensions
  - Definitions of the dimensions inherently ambiguous



- Explainability
  - Assessment results explainable to consumers
  - Results traceable to their root cause, to improve quality

- Efficiency
  - Assessment effort and time should be low

- Compliance
  - Fulfill organizational data governance processes
  - Comply to a legal framework, e.g., GDPR or the AI Act



- Adequacy
  - Is the data of sufficient quality or adequate for the task at hand?

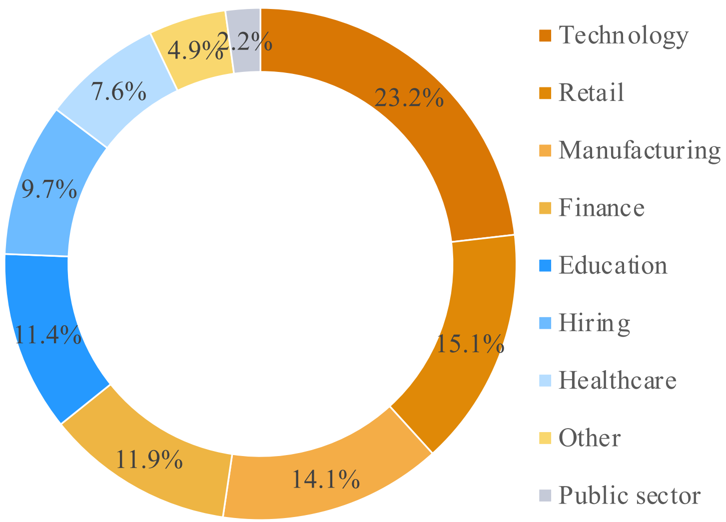
University of Amsterdam
Faculty of Science

Yichun Wang et. al. Machine learning practitioners' views on data quality in light of EU regulatory requirements: A European online survey. under review at JDIQ, 2025

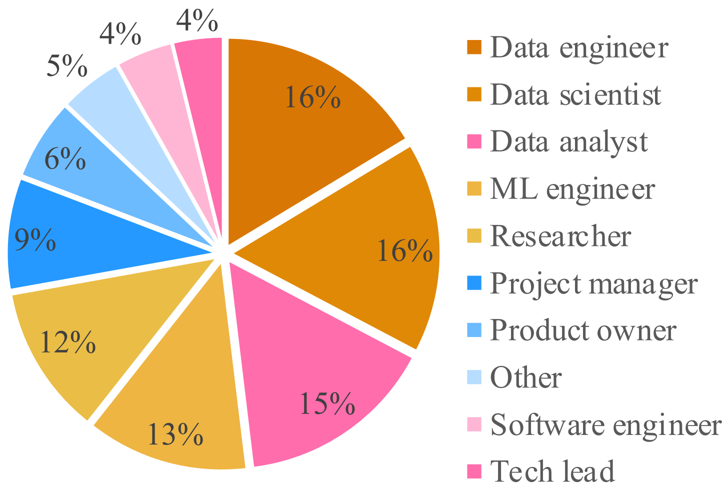
Data Quality and Compliance

Data management literature + EU regulatory requirements = Framework alignment

Industries



Roles



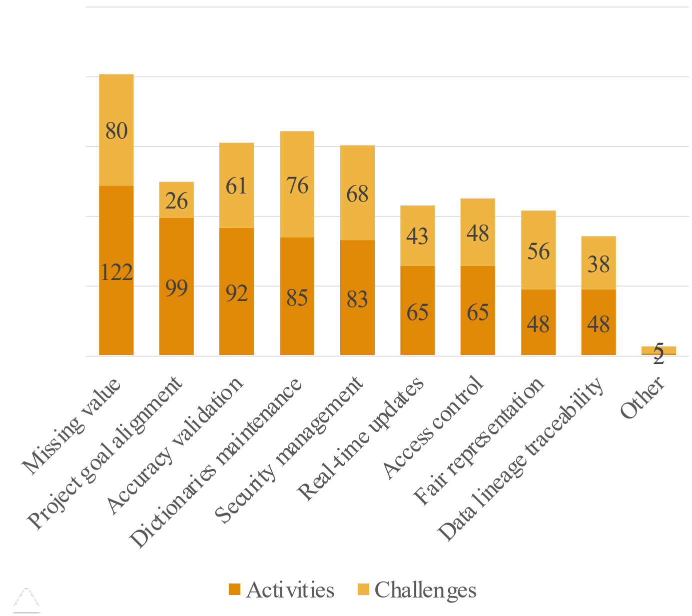
Survey overview

- Involving over **180** data practitioners across **24** EU countries
- Representing a diverse range of **domains** (inc. high-risk AI industries) and **roles**
- Empirical data on the **challenges** and **practices** related to data quality and regulatory compliance

UNIVERSITY OF AMSTERDAM
Faculty of Science

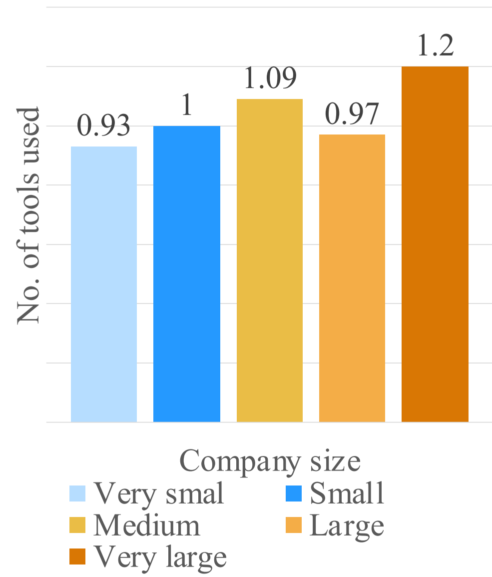
Data Quality and Compliance

Key DQ compliance challenges



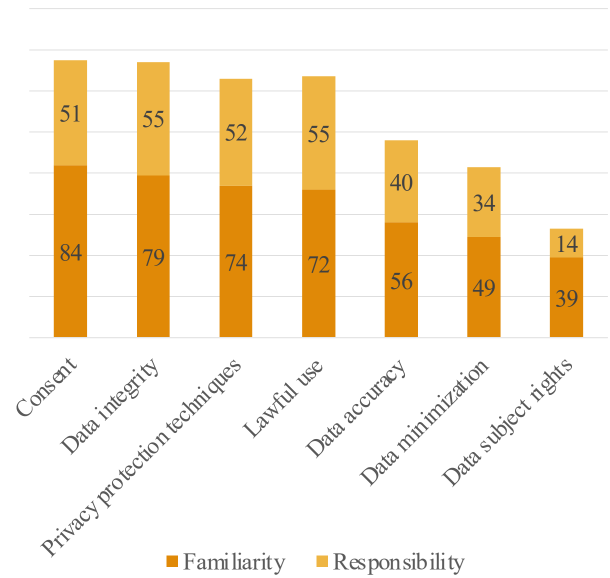
Despite active use of fairness protocols and privacy-enhancing technologies, these challenges persist:
Missing critical values (43%)
Lack of documentation (41%)
Privacy issues (37%)
Incorrect values (33%)
Data bias (30%)

Unmet needs & opportunities



Demand for more integrated tooling and clearer collaboration workflows between
ML and legal/compliance teams.
Most requested features:
Automated data validation (n=69; 37%)
Compliance frameworks (n=68; 37%)
Privacy protection mechanisms (n=58; 31%)

Familiarity & responsibility



Practitioners collaborating with legal teams report familiarity with nearly twice as many personal-data aspects as non-collaborators (3.81 vs. 1.94 aspects) and take responsibility for more (2.40 vs. 1.26)

Provenance based compliance system (Ongoing)

A data lineage-and-provenance system that is ”EU regulatory-aware” and “compliance-ready”(for the GDPR & AIA), while enforcing data-quality constraints

Actionable, collaborative, automated

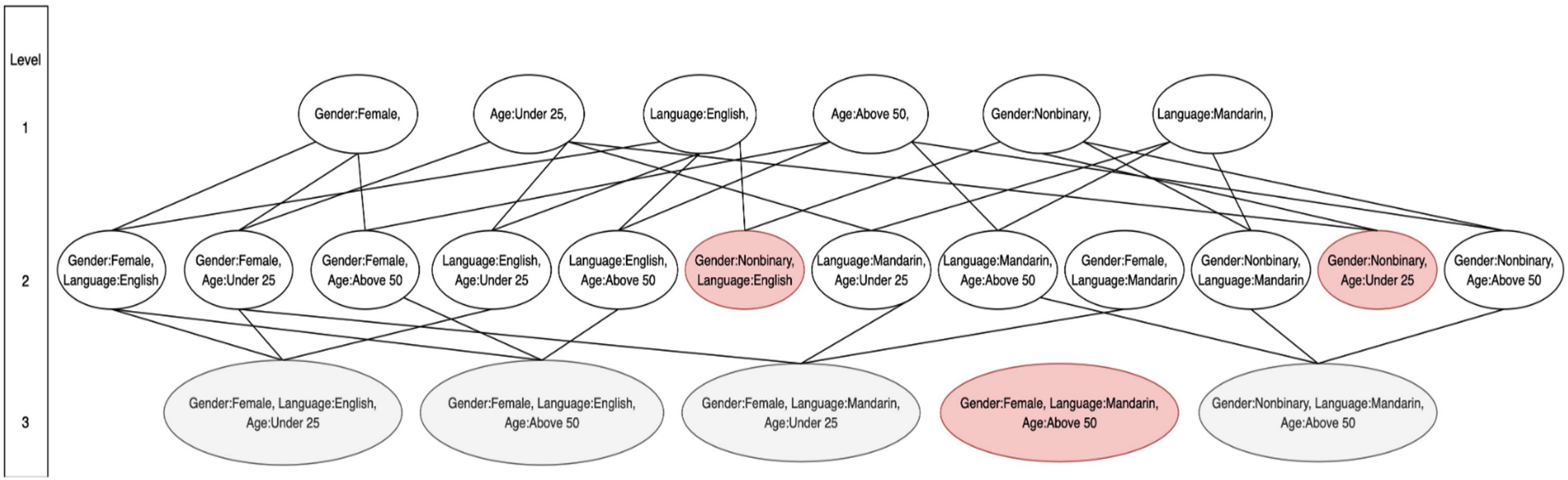
UNIVERSITY OF AMSTERDAM
Faculty of Science

A Catalog of 51 Data Errors – (Ongoing)

Error Manifestation	Error Name	Data Granularity						Context		Cause	
		Value	Tuple	Attribute	Relation	DB	Mult. DBs	Syntactic	Semantic	Action	Inaction
Missing	Missing Value [19, 46, 50, 58]	X						X		X	
	Disguised Missing Value [54, 57]	X		X					X	X	
	Partial-Empty Tuple/Attribute [50]		X	X					X	X	
	Missing Tuple [30, 46]		X						X	X	
	Empty Attribute [21]			X					X	X	
Incorrect	Invalid Value/Tuple [46, 58]	X	X						X		X
	Out-of-Vocabulary Word	X							X	X	
	Misspelling [19, 34, 50, 58]	X							X	X	
	Typo [11, 61]	X							X	X	
	Misscan [35, 48]	X							X	X	
	Incorrect Encoding [34]	X						X		X	
	Synonyms [45]	X							X	X	
	Word Transposition [34, 58]	X							X	X	
	Incorrect Unit [34]	X						X		X	
	Noise	X								X	
	Misfielded Values [58]		X						X	X	
	Contradiction [46]		X						X	X	
	Outlier [25]			X					X	X	
	Syntax Violation [19, 50]	X						X		X	
	Heterogeneous Formatting [30, 50]			X				X		X	
	Heterogeneous Unit [30, 50]			X				X		X	
	Incorrect Reference [30, 50]				X				X	X	
	Integrity Constraint Violation [30]				X				X	X	
	Domain Constraint Violation [19, 30, 50]			X					X	X	
	Uniqueness Violation [19, 50, 58]				X			X	X		
	Attribute Dependency Violation [58]			X					X	X	
	Functional Dependency (FD) Violation [19, 30, 50]				X				X	X	
	Conditional FD Violation [30]				X				X	X	
	Cyclic Dependency Violation [50]				X				X	X	
	Business Rule Violation [19, 50]				X				X	X	
	Database Administrator Rule Violation [19]					X			X	X	
	Legal Rule Violation [19]					X			X	X	
	Outdated Data [19]	X	X		X	X			X		X
Redundant	Duplicate Value [19]	X							X	X	
	Semantic Ambiguity [34, 50]		X						X	X	X
	Irrelevant Data [19]		X						X	X	
	Duplicate Tuples [19, 30, 46, 50, 58]		X						X	X	
	Duplicate Attributes			X					X	X	
	Biased Data				X				X	X	X
	Heterogeneous Schema					X	X	X		X	X

Table 1. Classification of data error based on how they manifest in data as main category, and three other possible classifications, granularity, context, and cause.

Measuring Diversity (Ongoing)



Gender	Language	Age
Female	English	Under 25
Nonbinary	Mandarin	Above 50
Female	Mandarin	Under 25
Female	English	Above 50
Female	English	Above 50
Nonbinary	Mandarin	Above 50

University of Amsterdam
Faculty of Science

Summary

- Data and Information Quality Research
- Data Quality and AI Systems
- Cleaning for ML
- Data Quality Assessment

Privacy
Uniqueness
Relevance
Understandability
Security
Accuracy
Added value
Consistent representation
Fairness
Reputation
Diversity
Representativity
Precision
Consistency
Reliability
Believability
Traceability
Completeness
Portability
Accessibility
Documentation
Balance
Timeliness
Amount
Transparency
Recoverability
Efficiency
Compliance
Representativity
Ease of manipulation