The Cognitive Data Wizardry for Ultimate Robustness and Transparency

Dr Maryam Farsi
Senior Lecturer in Engineering Optimisation
Centre for Digital Engineering and Manufacturing | Cranfield University
September 2024

CIRP Annals - Manufacturing Technology 72 (2023) 385–388

Contents lists available at ScienceDirect

CIRP Annals - Manufacturing Technology

journal homepage: https://www.editorialmanager.com/CIRP/default.aspx

ELSEVIER

Cognitive data imputation: Case study in maintenance cost estimation

John Ahmet Erkoyuncu (2)[a,*], Bernadin Namoano[a], Dominik Kozjek[b], Rok Vrabič (2)[b]

[a] Cranfield University, School of Aerospace, Transport and Manufacturin

[b] University of Ljubljana, Faculty of Mechanical Engineering, Slovenia

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

ELSEVIER

Civil aircraft engine operation life resilient monitoring via usage trajectory mapping on the reliability contour[☆]

Hang Zhou [a,b,*], Maryam Farsi [c], Andrew Harrison [d], Ajith Kumar Parlikad [b], Alexandra Brintrup [b]

[a] James Watt School of Engineering, University of Glasgow, UK

[b] Institute for Manufacturing, Department of Engineering, University of Cambridge, UK

[c] School of Aerospace, Transport and Manufacturing (SATM), Cranfield University, UK

[d] Rolls-Royce plc, Derby, UK

Context

Cost Estimation

Cost data/ information

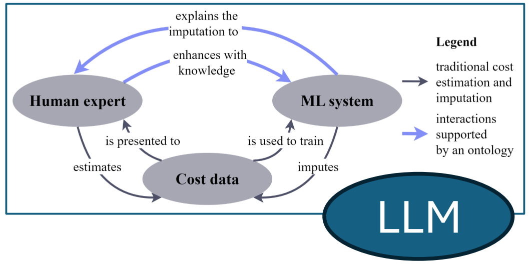
Historical data

Knowledge

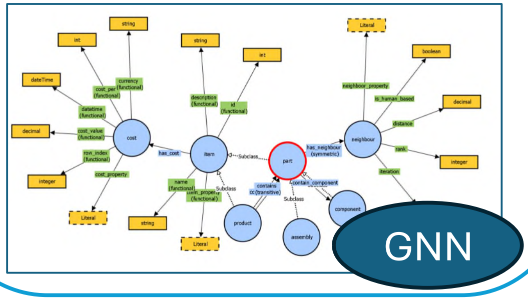
Quality challenge

Cognitive Data Imputation

Manages uncertainties in low-quality data, information and knowledge by learning from trends and contexts



Ontology – Facilitates interaction between knowledge and the machine learning system

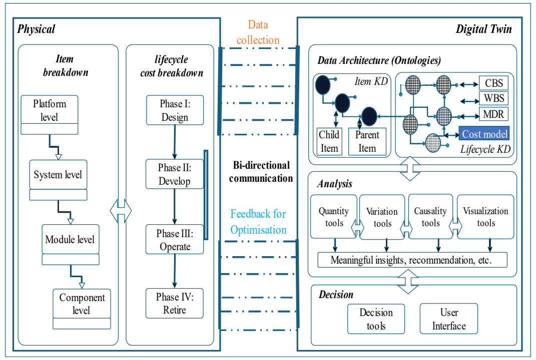


Manages uncertainties

Large Language Model (LLM)
Graph Neural Network (GNN)

Digital Twinning (models, shadows, twins)

Creates virtual replicas of physical assets and processes



Real-time/or not monitoring and analysis
Predictive modelling and enhanced the robustness in analyses over time

Robust Decision Making

Data Quality Challenges

Incomplete data
Missing data and meta data, unrecorded data, accessibility.

Inaccurate data
Incorrect cost values, currencies, use, outdated

Duplicate data
Poor data integration, lack of unique ID for cost items

Inconsistent data
Variation and lack of standardisation in data formatting, data management

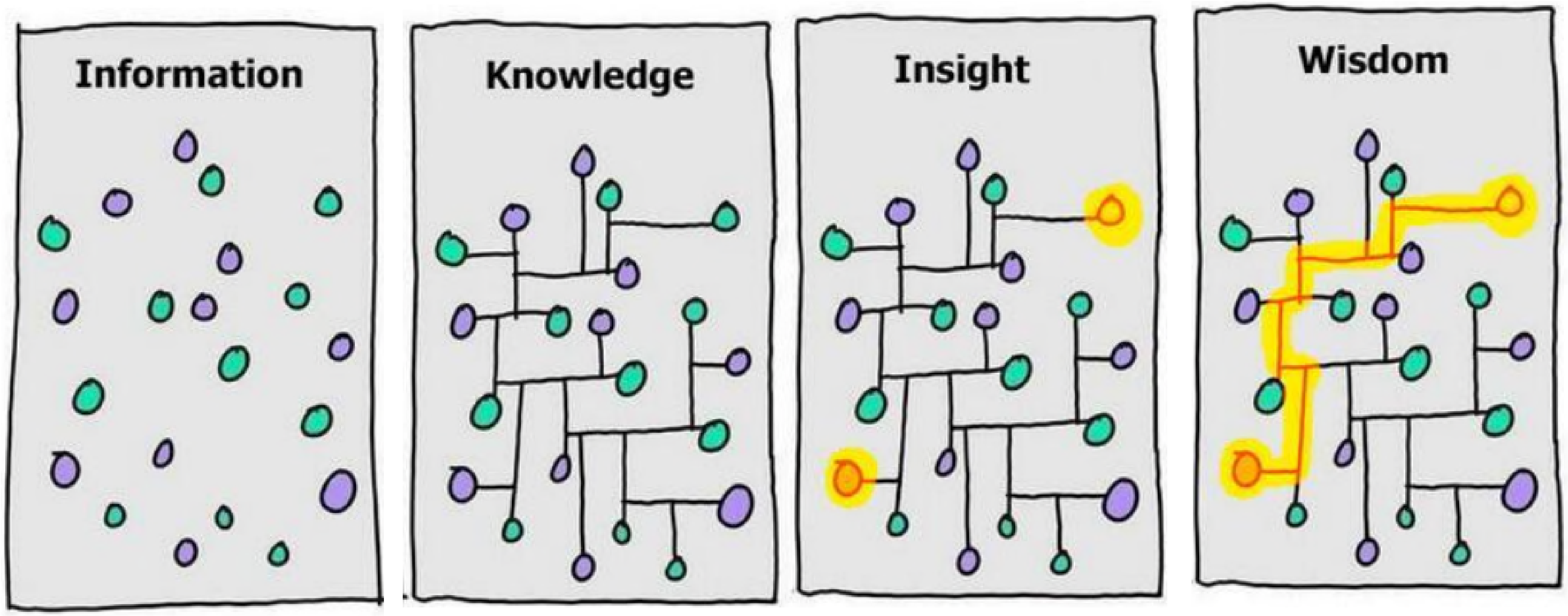
Lack of standardisation
Terminologies, cost metrics, cost measures, currencies, data collection

Data Fragmentation
Multiple sources, lack of integration – siloed systems

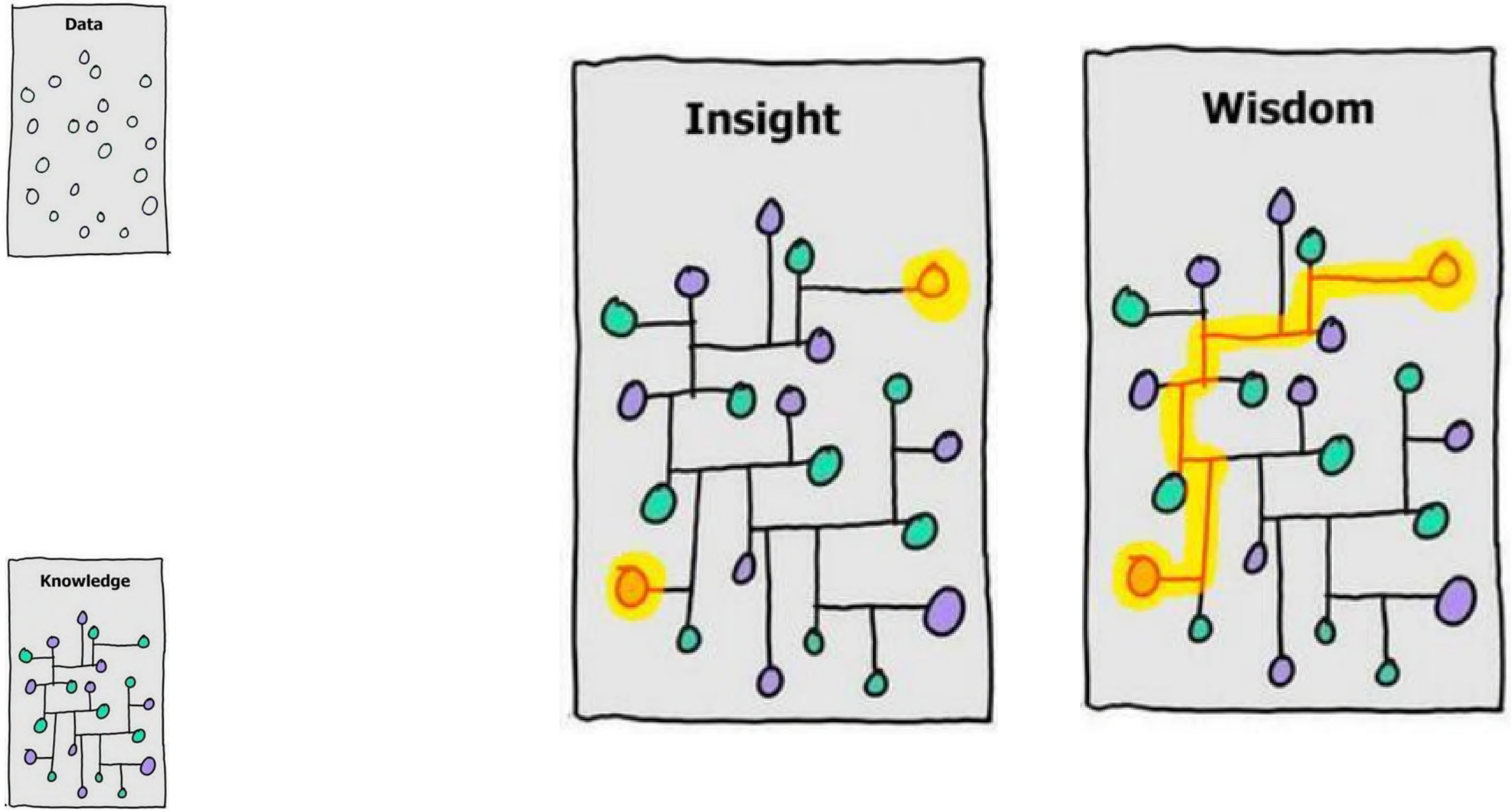
Unclear data ownership
Lack of accountability, weak data governance policies, lack of data quality control measures

Unreliability in data
Lack of trust, manipulation or corruption.

Information Continuum



Information Continuum in Cost estimation



Information/Knowledge Quality Challenges

Incomplete Knowledge Capture
Lack of expert knowledge, poor knowledge sharing

Knowledge Fragmentation
Knowledge exists across different teams, lack of integration and consolidation

Obsolete Knowledge
Outdated over time due to changes in market conditions, New technologies, etc.

Knowledge Loss
Key employees with critical cost-related knowledge may leave

Bias in Knowledge
Subjectivity, Personal or organisational biases

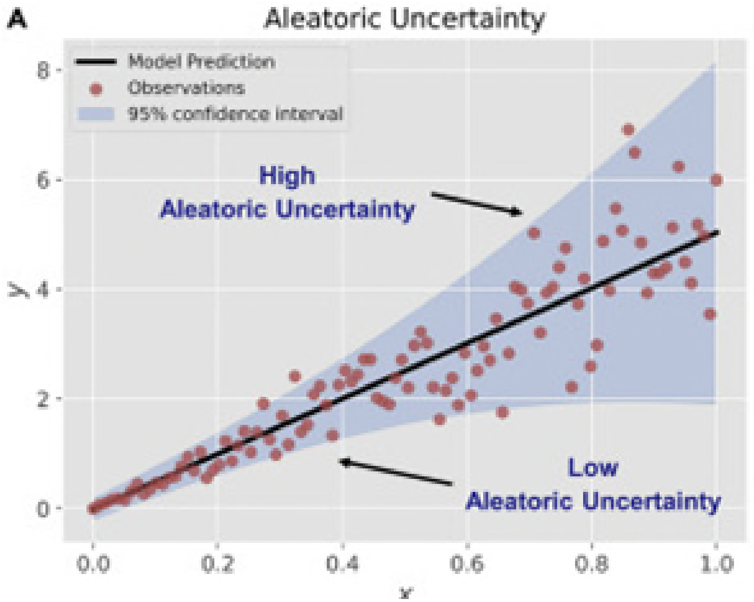
Imprecision
lack of accuracy, precision, or specificity in knowledge capturing, vagueness, ambiguity

Overreliance
Rely too heavily on internal historical data, Lack of External Benchmarking

Poor data sharing
Asymmetric information between organisations, lack of alignment between data sharers

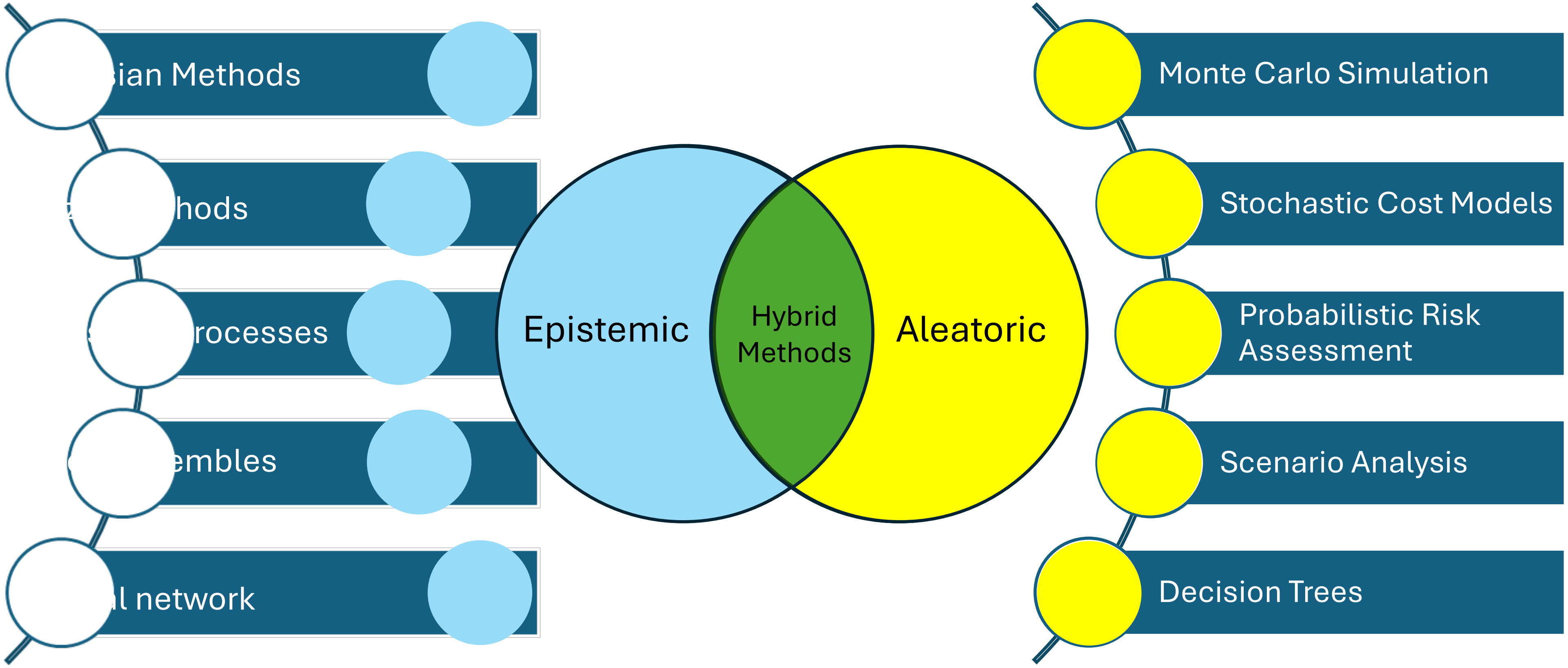
Predictive Uncertainty: aleatoric vs epistemic

- Data-related quality issues generally cause ‘aleatoric’ uncertainty
- Information and knowledge-related quality issues generally cause ‘epistemic’ uncertainty



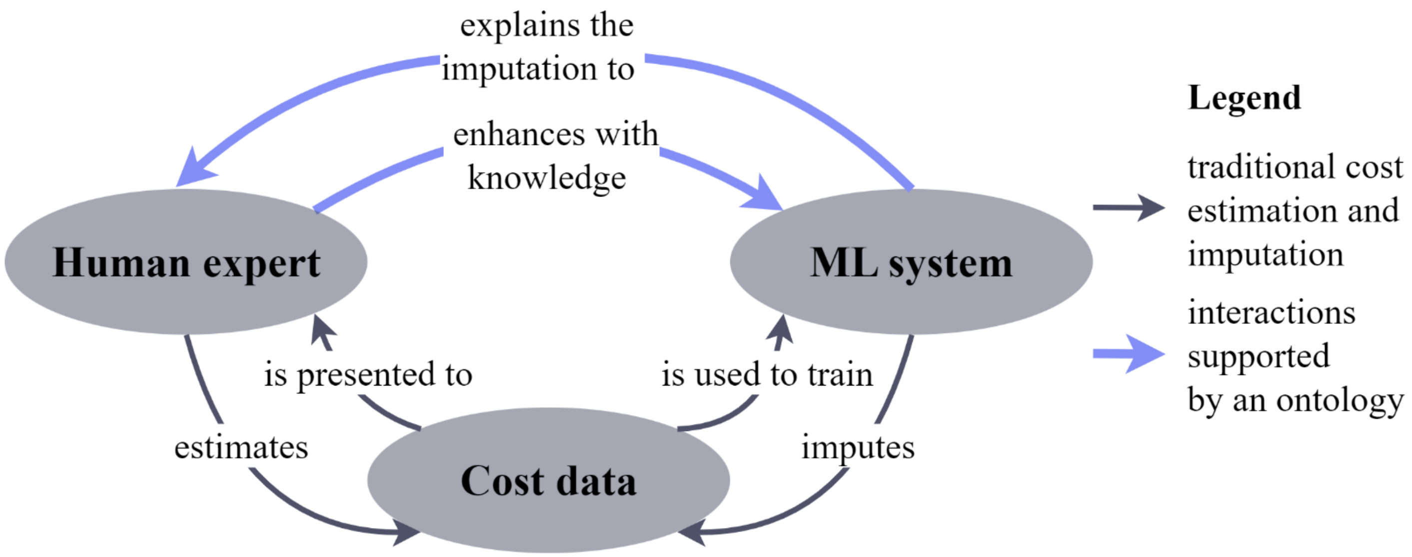
Yu, J., Wang, D. and Zheng, M., 2022. Uncertainty quantification: Can we trust artificial intelligence in drug discovery?. *Iscience*, *25*(8).

How do we capture and quantify uncertainty in cost estimation?



Explanatory Interactive Machine Learning (XIL)

Learning is a co-adaptive process in which human cognitive abilities are used to modify the behaviour of an ML system (and vice-versa).

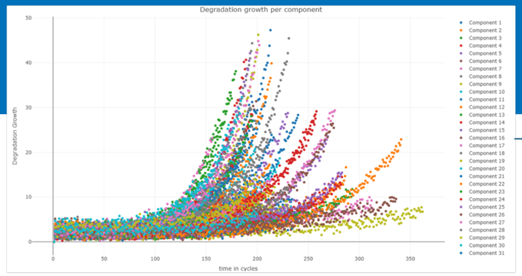


Relationships between the human, the ML system, and the data

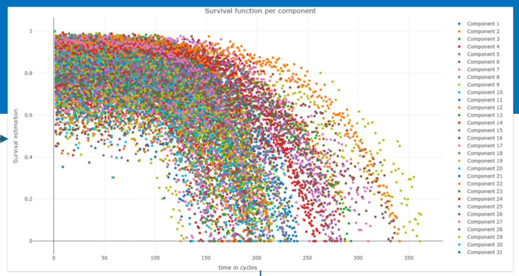
Erkoyuncu, J. A., Namoano, B., & Kozjek, D. (2023). Cognitive data imputation: Case study in maintenance cost estimation. *CIRP annals*, *72*(1), 385-388.

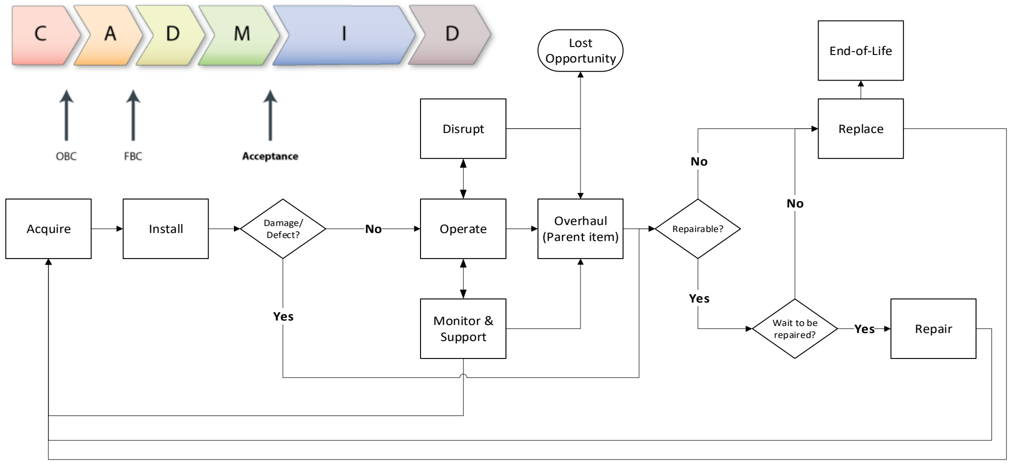
Reliability informed Cost

Degradation growth based on components' lifedata

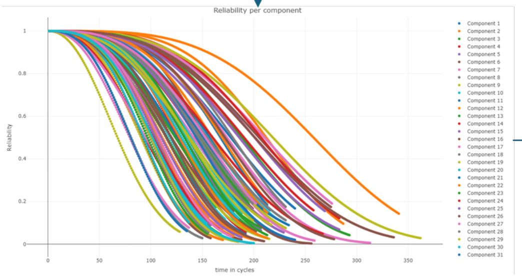


Survival analysis

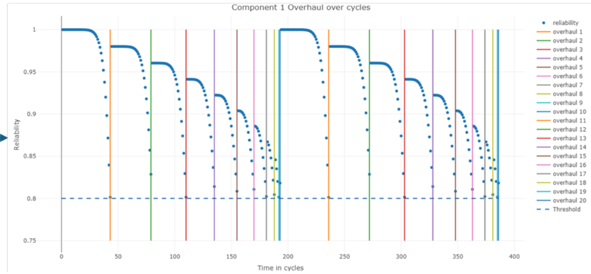




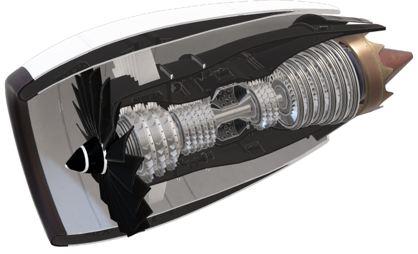
Reliability model



Predicted number of overhauls caused by the component over the operating time



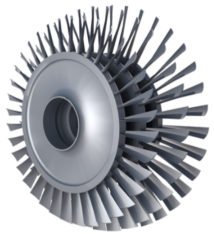
System level



LPC	IPC	HPC	CBT
LPT	IPT	HPT	EGB

HPT

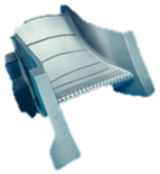
Module level



HPT rotor	NGV

NGV

Component level



Deterioration Mechanism

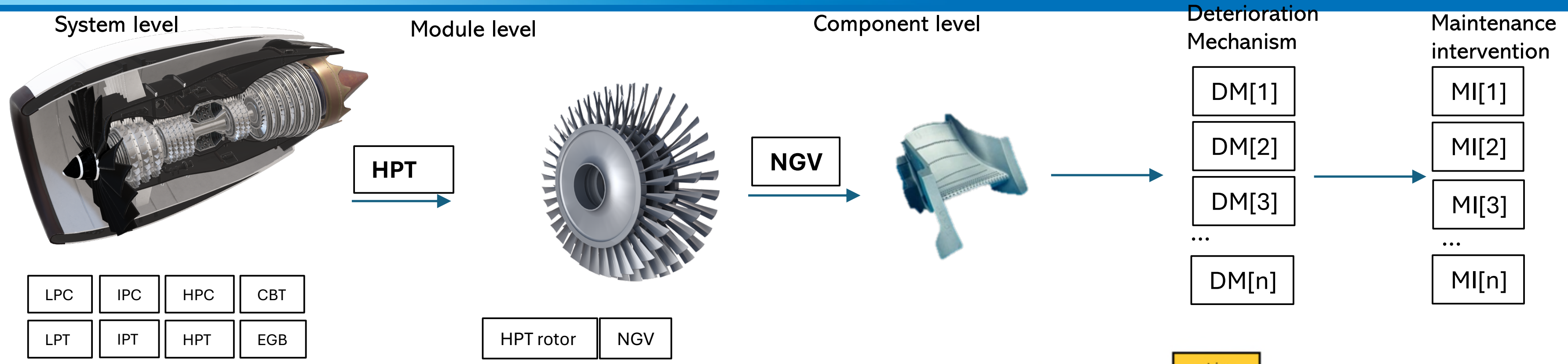
- DM[1]
- DM[2]
- DM[3]
- …
- DM[n]

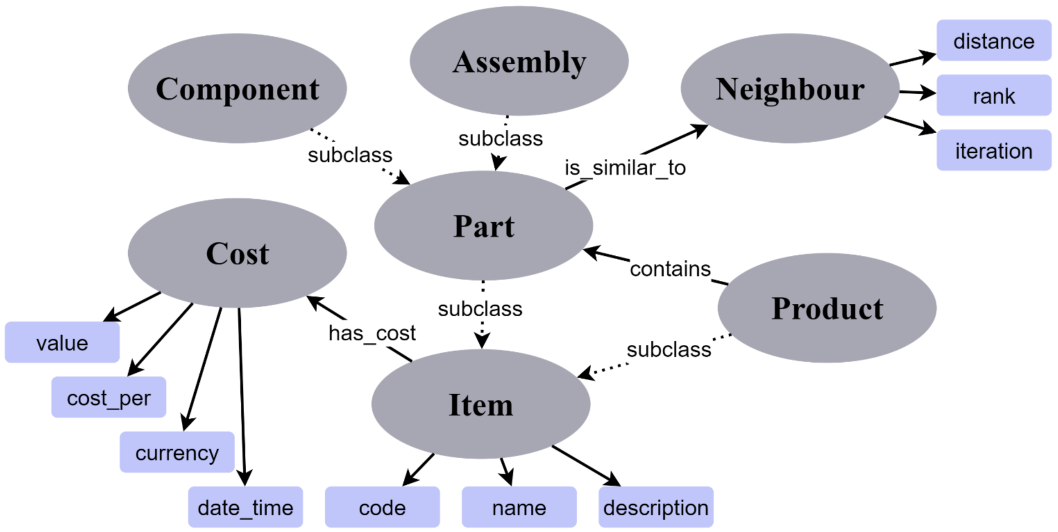
Maintenance intervention

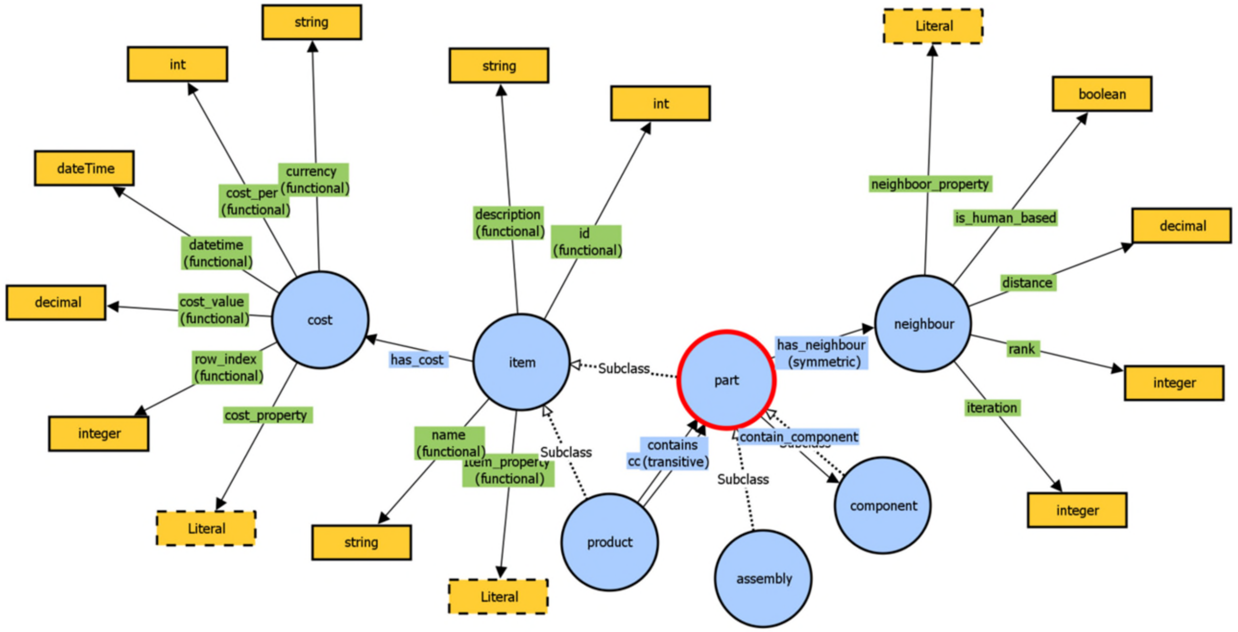
- MI[1]
- MI[2]
- MI[3]
- …
- MI[n]

Zhou, H., Farsi, M., Harrison, A., Parlikad, A.K. and Brintrup, A., 2023. Civil aircraft engine operation life resilient monitoring via usage trajectory mapping on the reliability contour. *Reliability Engineering & System Safety*, *230*, p.108878.

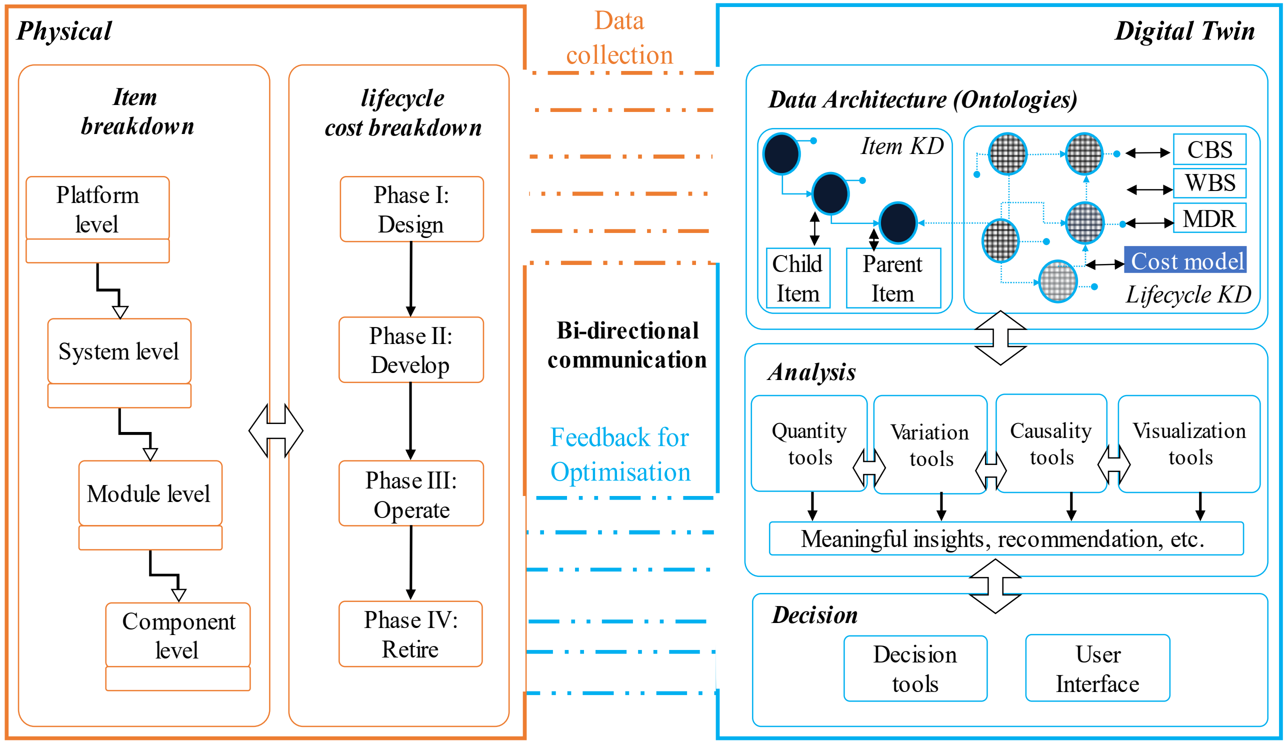
Cost Data – Ontology Model & Knowledge Graph

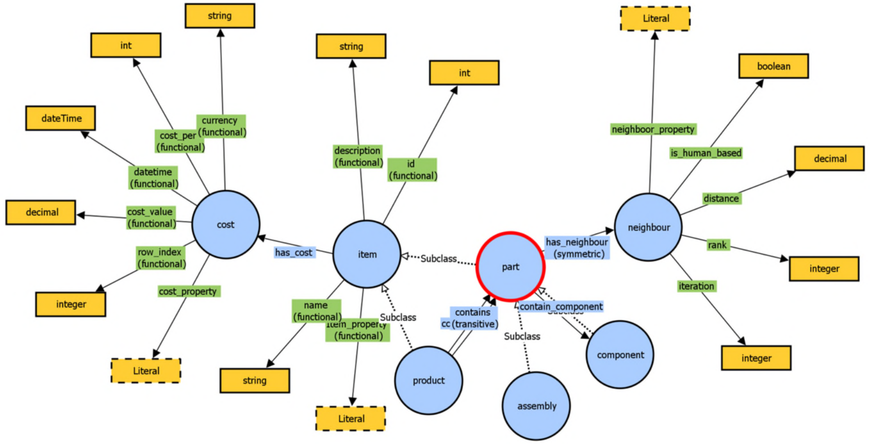


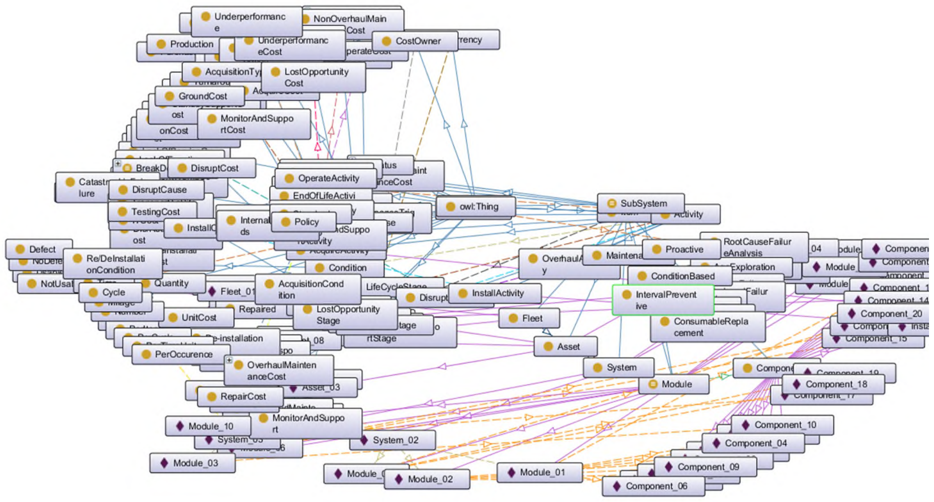




Ontology-based digital twin

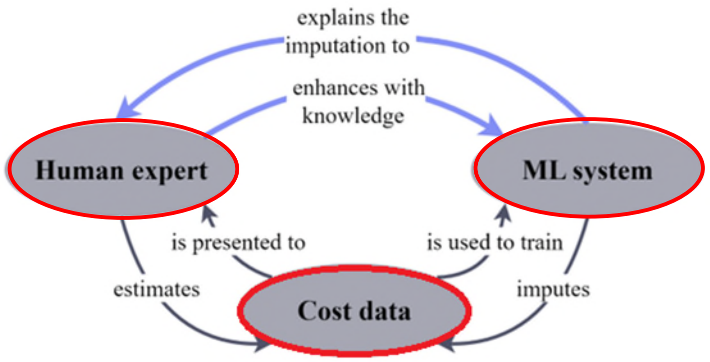






Farsi, M., Erkoyuncu, J.A. and Harrison, A., 2020. A Super Simple Life-cycle Cost Estimation Model with Minimum Data Requirement. Available at SSRN 3718042.

Data Imputation
Graph-based Neural Networks



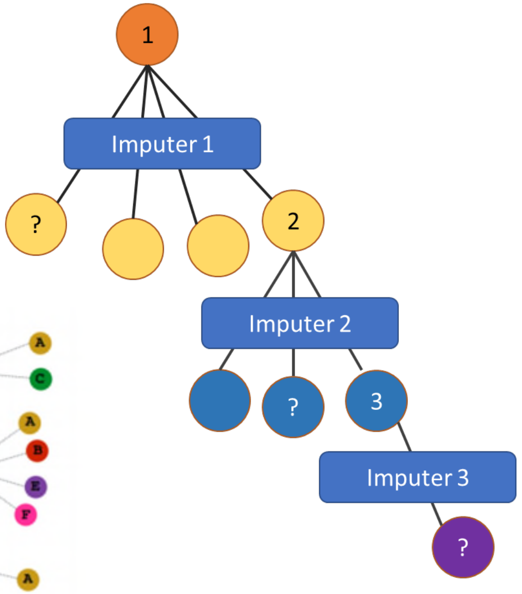
k-nearest neighbours (kNN) selected as the imputation method

- ➢ Interpretability, explainability
- ➢ Ability to incorporate user feedback by modifying the distance metric ***d***

- Imputed cost ($\hat{c}_i$):

  c_j: neighbouring components
  $d_{i,j}$: the distances between them

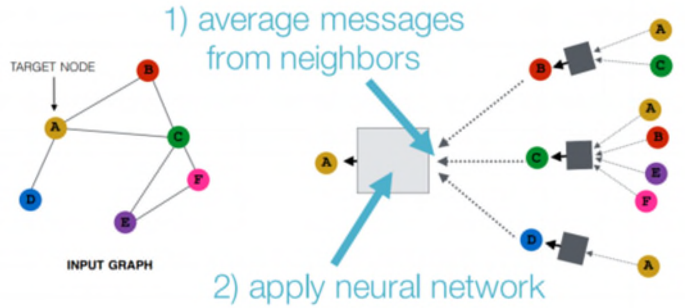
$$\hat{c}_i = \frac{\sum_{j=1}^{k} \frac{c_j}{d_{i,j}}}{\sum_{j=1}^{k} \frac{1}{d_{i,j}}}$$



- Evaluation (ability of imputation) using R^2 metric:

  $\bar{c}$: the average of the costs of all components

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(c_i - \hat{c}_i)^2}{\sum_{i=1}^{N}(c_i - \bar{c})^2}$$



Erkoyuncu, J. A., Namoano, B., & Kozjek, D. (2023). Cognitive data imputation: Case study in maintenance cost estimation. *CIRP annals*, *72*(1), 385-388.

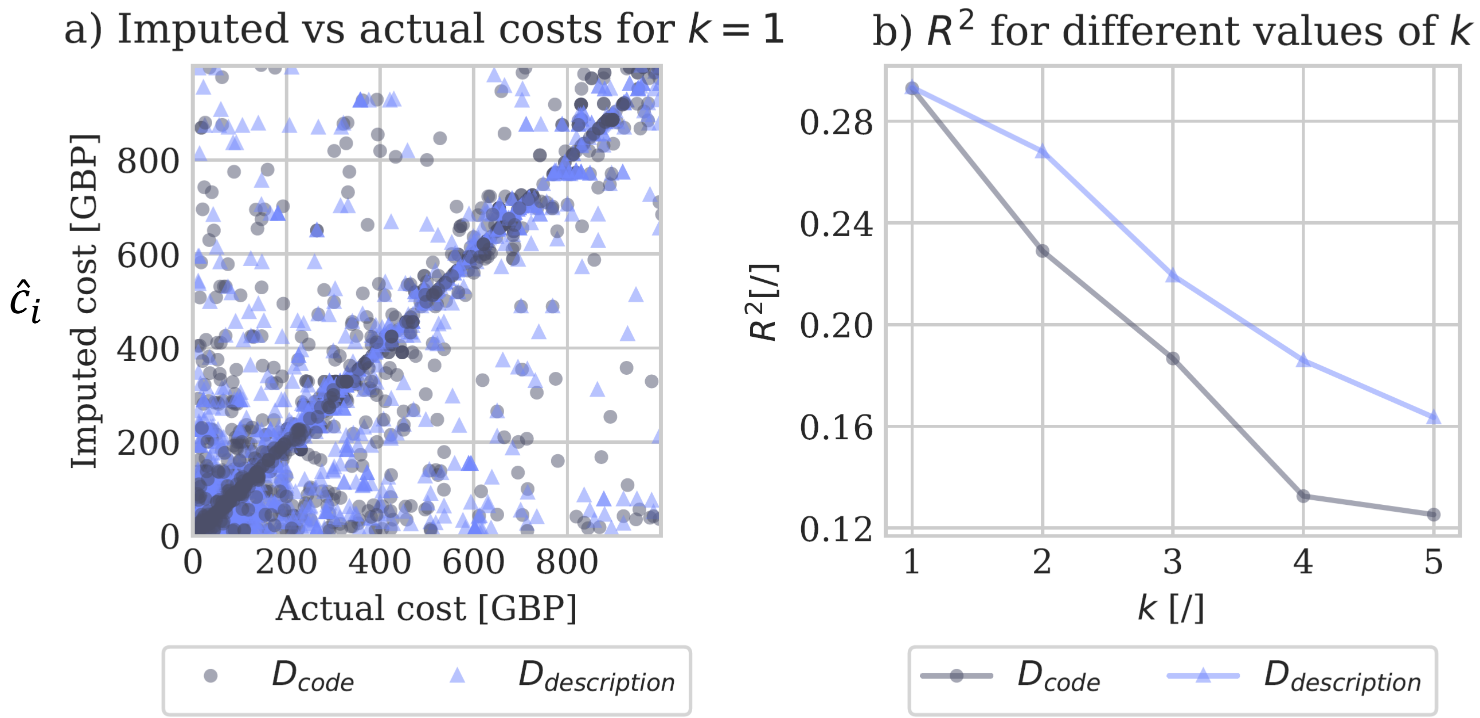
Case Study
Maintenance Data from POs

- **Company:** Leading provider of complex engineering services in marine and aerospace sectors across Europe.
- **Business Focus:** Managing high-value assets (e.g., ships, submarines, helicopters) with a strong emphasis on asset management.
- **Key Goal:** Improve asset reliability, availability, maintainability, and safety, minimise operating costs, and reduce time between failures.
- **Data Focus:**
  - 14,000 purchase orders related to aerospace asset spare parts.
  - 70 attributes covering maintenance, operations, and financial aspects.
- **Data Challenges:** missing values, formatting issues, typos, ambiguous, redundant and inaccurate data, outliers, data out of range

Erkoyuncu, J. A., Namoano, B., & Kozjek, D. (2023). Cognitive data imputation: Case study in maintenance cost estimation. *CIRP annals*, *72*(1), 385-388.

Case study - Imputation method analysis

- Component codes and descriptions as initial metrics of similarity.
- To be improved via XIL, ontology.



An analysis of **kNN imputation** using the **Levenshtein** distance of part codes and part descriptions as the similarity metric

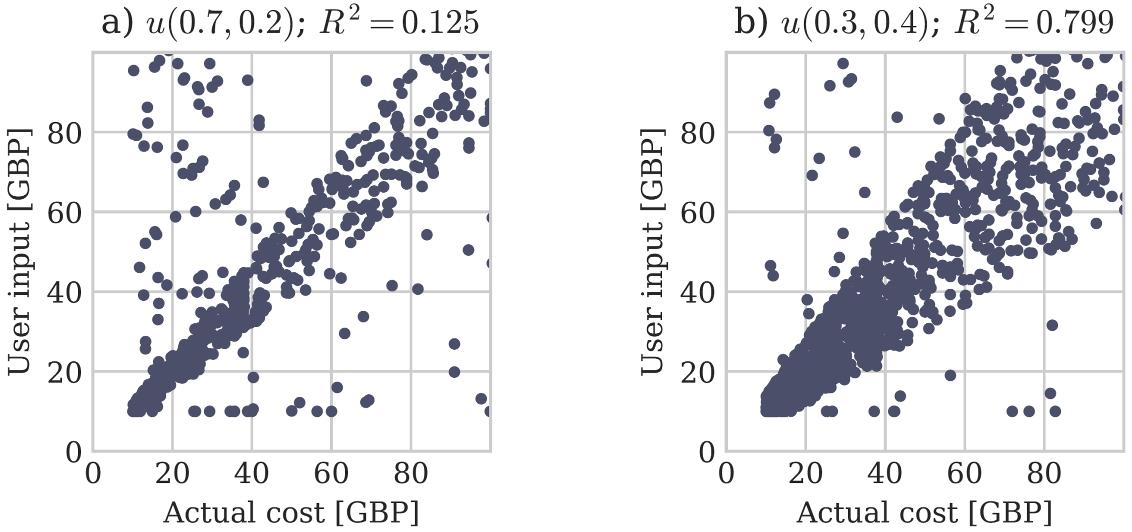
Pearson correlation and **Goodman tau** were computed respectively for quantitative and qualitative variables of the dataset.

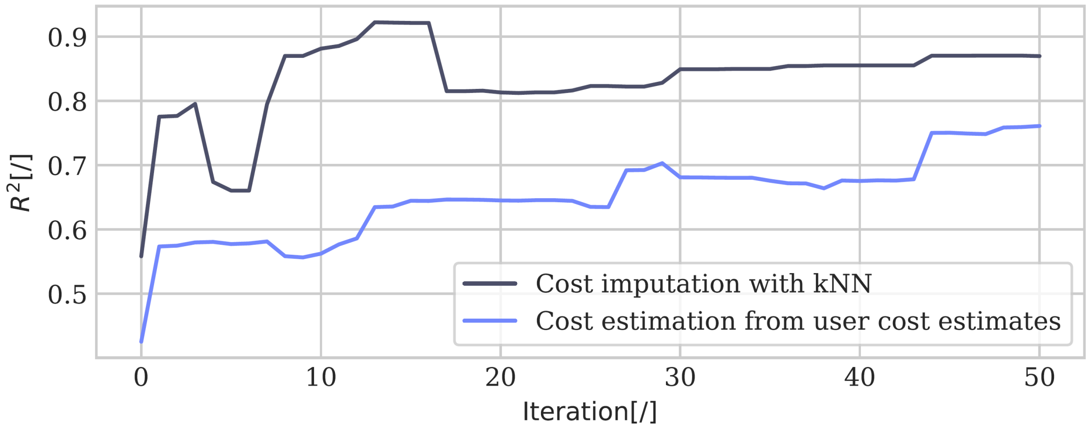
Case study - Simulation Study

Simulation study with artificially generated users: The main goals are to evaluate the possibility of improving the system through interaction and to estimate how much interaction is required for the improvement.

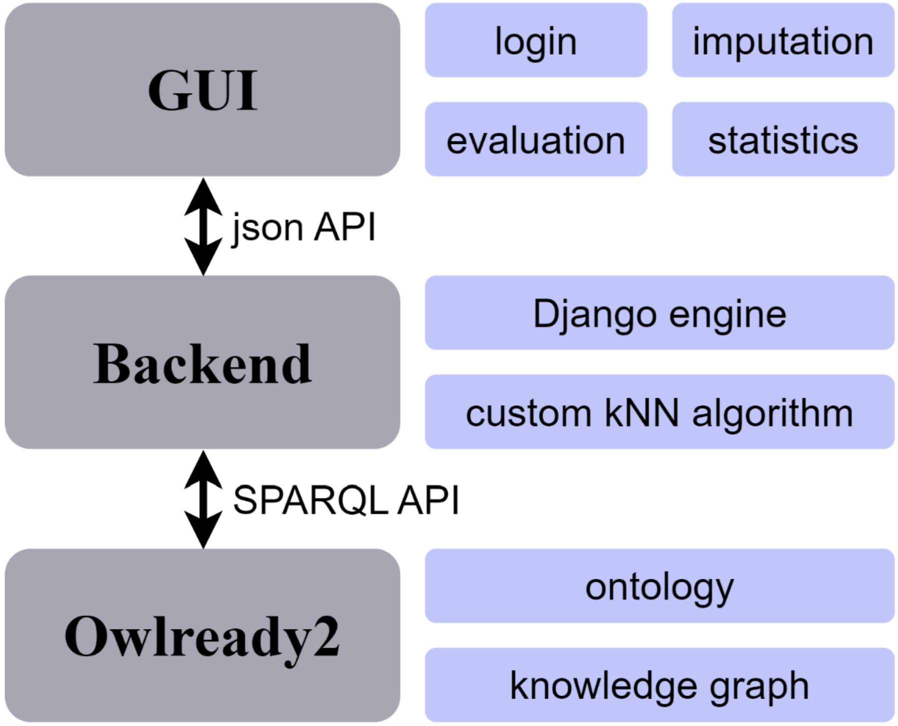
100 components are selected at random and simulated to have missing cost values. In each iteration, a new, randomly generated user is given 10 components for which they determine the cost and list similar components.

Simulated user with random inputs $u(\pi, \sigma)$ where $(1 - \pi)$ is the probability that the user cost setting is sampled from a normal distribution around the actual cost with σ as the standard deviation of that normal distribution.

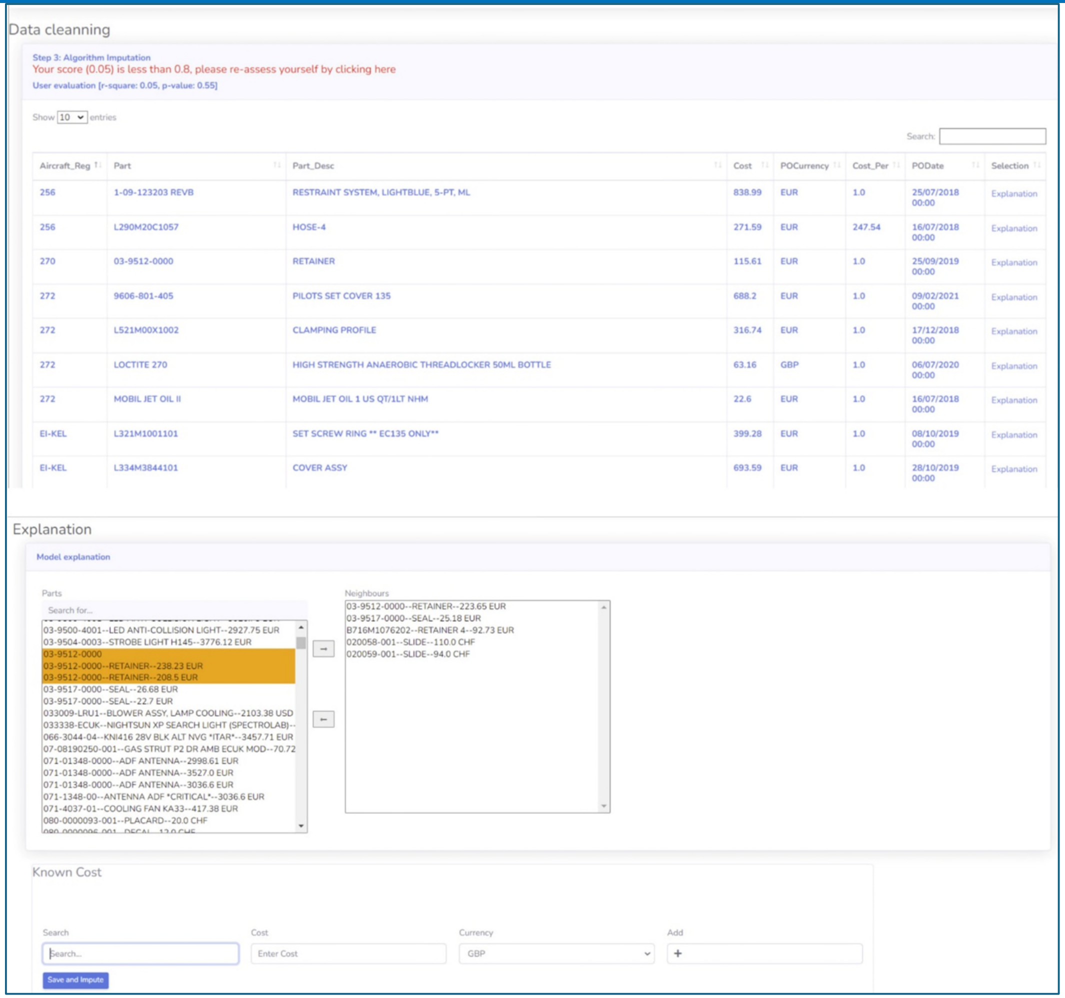




Case study - User Study



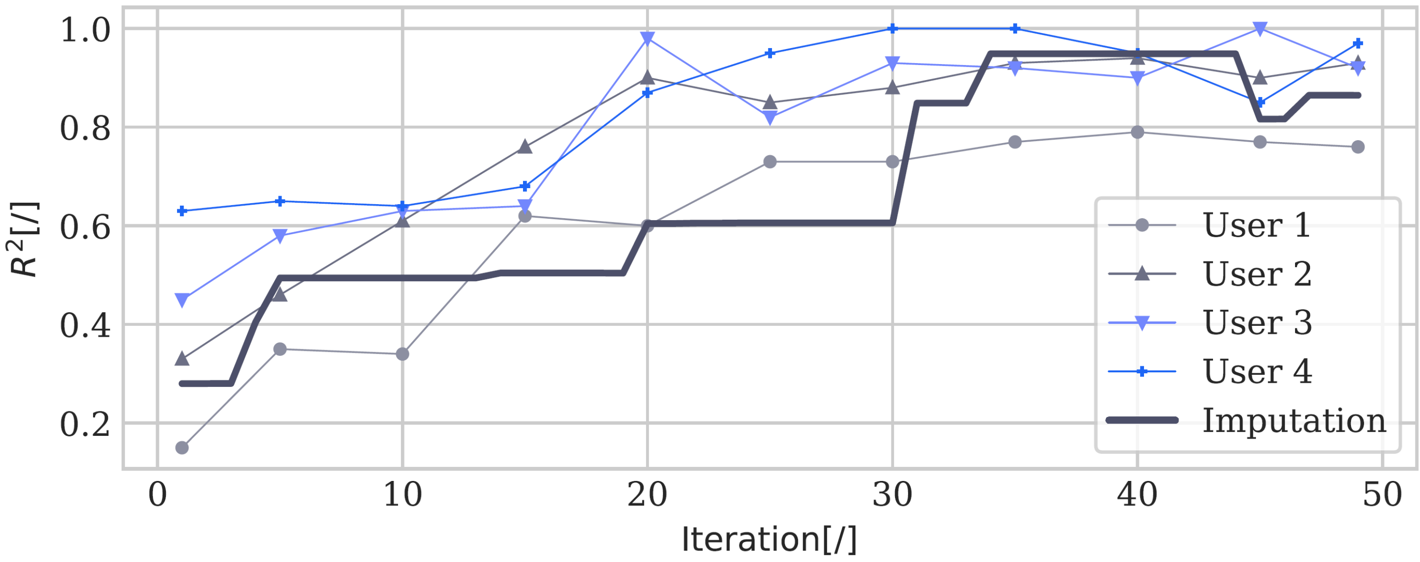
An architecture for Python web-page application



Examples of the GUI.

Case study - User Study

- Four real users: 2 experts (users 3 and 4) and 2 non-experts with limited knowledge (users 1 and 2)
- The scores show an increasing trend over time, indicating an improvement in the estimates and the underlying human understanding of the data.



An example of four users' cost estimation showing both user and algorithm improvement.

Concluding remarks

- **Focus:** Tackles both aleatoric and epistemic data uncertainty challenges in cost estimation.
- **Traditional Approaches:** Previous methods relied heavily on existing data, missing the benefits of expert human input for improving estimates.
- **Case Study:** Applied to an aerospace maintenance case with 14,000 procurement-related data entries to address data uncertainties in cost estimation.
- **Approach:** Cognitive data imputation using Explanatory Interactive machine learning
- **Results:** the results confirmed complementary improvements in cost estimation, integrating user expertise with machine learning (ML) based imputation techniques.

Thank you

Dr Maryam Farsi

maryam.farsi@cranfield.ac.uk