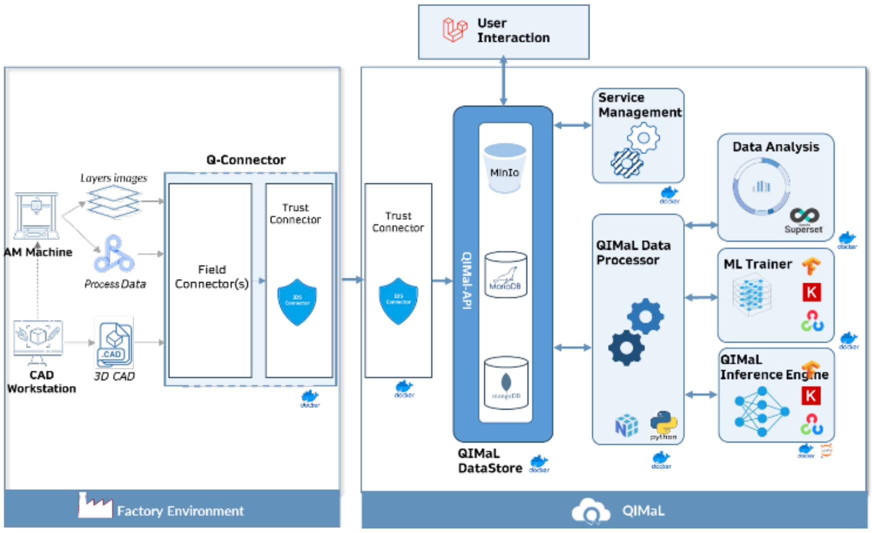
Technical Specification

NAME OF THE COMPANY: Youbiquo
CHALLENGE: OPTIMIZATION OF VISUAL INSPECTION PROCESS
TRACK: TRACK 1

1. **TECHNICAL SCOPE:** Summarize the solution developed during the EXPERIMENT phase: how have you finally addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

QIMaL aims to optimize the visual inspection process in additive manufacturing (AM) by identifying and preventing defects during the production of metal parts using SLM or PBF. Defects caused by process randomness and laser powder fusion irregularities can lead to 3D print failures and wasted time, resources, and materials. To overcome this, QIMaL combines data from various sources, such as machines, CAD design and HD images, to identify and prevent defects. The entire solution, along with its context, is depicted in the diagram. The context – Factory Environment – is made up of the set of workstations and AM machines that participate in the design and manufacturing processes. These processes return digital products made up of 3D CAD, Process Data, Images of the printed layers. These data are made available to QIMaL through the Q-Connector. The Q-Connector includes sw connectors to workstations and machines and the data provider of the IDS Connector. Sw connectors can implement OPC UA reference standard. IDS Connector enables a trusted data exchange towards QIMaL where the data consumer of the connector is installed. Through API data are annotated and recorded in the DataStore. QIMaL DataStore manages three different data structures: MongoDB (nosql database), MariaDB (rdbms) and MinIO as object store. The rdbms is used to record transactional data about clients and service management (Service Management module), MongoDB is used to store data collections from the machines and the object store stores images and drawings. Data from the datastore are processed on request of ML Trainer Module, QIMaL Inference Engine and Data Analysis Module. A Convolutional Neural Network (CNN) model is used in ML Trainer Module to extract features from the overlap of images extracted from the CAD layer-by-layer and those obtained through video acquisition systems to predict the defect score for each of the defect sensors installed onboard the AM machine, obtaining the output representation (C). A Recurrent Neural Network (RNN) with a sliding time window is used to analyse process data, performing the same task, resulting in a second output representation called (P). Since the AM machine calculates defect scores through defect sensors, the predicted scores output by trained ML models are compared to actual scores using distance metrics, such as the Mean Absolute Error (MAE). This allows the Inference Engine to employ both models to select the defect source with a higher correlation with the values measured by sensors between layer images (C) and process data (P). Additionally, analysing inputs and outputs to the models through mathematical and statistical techniques allows to identify machine parameters or the CAD section that may have contributed to the defect, thereby facilitating the determination of corrective actions for the production or design phase. The result of the analysis produces a Defect Report which is stored in the DataStore; it reports the detected errors and the probable cause. Defects are mapped on the image of the printed layer and on the 3D CAD allowing its immediate visualization on the object. The Data Analysis module enables the processing of data from MongoDB using business analysis tools like Apache Superset. Superset's semantic layer is leveraged to create dashboards for machine parameter analysis and statistical charts for process control.



2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools finally selected to accomplish the challenge/Theme Challenges. Summarize the main results that you have obtained during the EXPERIMENT phase: data, insights, conclusions and the main contributions to solve the challenge/Theme Challenges.

Two distinct ML models are trained using TensorFlow, Keras, and OpenCV in a Jupyter environment. The CAD model is a CNN designed for processing images, while the Process Data Model is intended for handling data measured during printing. GPUs are employed to facilitate efficient parallel processing during both training and model deployment. The CNN model is trained in an unsupervised manner on the regression task of predicting layers defect scores on images (C). Since 3D CADs are sliced before the printing process and the slices can be overlapped on images taken by video systems, defects are related to the layers and can be analysed individually. Therefore, CAD and layer images are processed together through convolutional layers for feature extraction. The concept of Spatial Invariance is introduced through pooling layers in image processing to allow for the detection of key features of defects in different positions of the input. The extracted features are processed for the prediction of defect scores, using the Mean Absolute Error (MAE) or the Huber loss during training to measure the distance with the real scores outputted by defects sensors.

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981

BIG DATA VALUE This project is part of BDV PPP

CORE PARTNERS

DATA PROVIDERS

The training is conducted using adaptive gradient descent techniques, specifically employing either the Adam or RMSProp optimizer to enhance model performance. For process data, the temporal component is evident: a defect may be due to parameter variations in the previous moments or continued over time. The number of past time steps (time window) is a hyperparameter which can be estimated from the data. A RNN such as Long-Short-Term Memory (LSTM) and/or Gated Recurrent Units (GRU) will be used, capable of handling long-term dependencies. The model is trained on the same regression task to predict defect scores (P) using the MAE or Huber loss and Adam or RMSprop optimizer. In both cases, regularization techniques such as L1/L2 regularization, Dropout, EarlyStopping, Data Augmentation were adopted to mitigate overfitting of the training data. The learned representations are used together to determine the correlation of defects with 3D CAD slices and/or process data through the calculation of distance metrics such as MAE. Feature importance and gradient-based techniques are adopted to identify the parameters or portions of CAD that caused the errors, to prevent and mitigate recurring errors over time.
The numerical data acquired by QIMaL undergoes the application of methods and algorithms typical of Statistical Process Control (SPC). This is done with the aim of generating statistical charts, such as p-charts, x-bar & s charts, to facilitate the analysis of production trends.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Explain how the solution copes with the challenge/Theme Challenges requirements and how can it be adapted to other similar problems. What work is still pending to create a real/stable product if any? What TRL level is it in? Discuss how your solution could be integrated within multi-stakeholder data value chains.

QIMAL addresses the challenge posed because it enables the identification of the causes of production defects. QIMaL empowers manufacturers to pinpoint the root causes of defects in products manufactured through additive processes, effectively tackling this challenge by enhancing production quality. To achieve product maturity and stability, it will be necessary to expand the base of AM machine models that QIMAL can interface with. The key lies in the machine learning model created, which needs to be trained across a wide range of machine parameters. The TRL is rated at level 5. The integration of QIMAL within a multi-stakeholder data value chain does not pose specific technological challenges. Reference to communication standards with the field environment, such as OPC UA, and the use of the IDS connector, ensure a secure and seamless integration of stakeholders using additive manufacturing. The results of the analyses conducted on the acquired data can be utilized by other stakeholders, such as machine manufacturers, research centers, or end-users.

4. **DATA GOVERNANCE AND LEGAL COMPLIANCE:** Describe the security level of the solution, i.e. how authentication, authorization policies, encryption or other approaches are used to keep data secure. Explain how the solution is compliant with the current data legislations concerning security and privacy (e.g. GDPR). Explain how you tackle data governance and sovereignty aspects in your solution.

In terms of data management, it is crucial to emphasize that even though the system will not handle personal data, the project will prioritize the principles of privacy and confidentiality. This means that each customer will have exclusive access to their own data, and the project will adhere to international rules and standards such as GDPR, ISO 27017, and ISO 27001 to ensure that data is handled appropriately. To address key challenges related to data interoperability, trust, and security in the data value chain an instance of the IDS Connector is deployed in the QIMaL infrastructure. This instance is based on the OCI image provided by the International Data Spaces Association and plays a crucial role in the data ingestion process, which can be summarized as follows: data connectors, on both the provider and consumer sides, establish a contractual agreement; the QIMaL-connector, acting as a consumer, receives an artifact URL as part of this agreement; the data provider or owner – in the factory environment - configures a data ingestion service using a web GUI; subsequently, the artifact URL is sent to the ingestion endpoint, which processes and stores the received content via the connector, following the service configuration.
Access to secured data are be enabled through a protected administration interface, which utilizes SSL encryption via the https protocol To ensure data access is restricted to authorized users and maintained securely, APIs with tokens and authentication keys are utilized. This ensures data access is limited to authorized personnel and safeguards data privacy and security. In addition, the APIs will incorporate an application-level encryption system to ensure data stored is encrypted."

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process followed for the final product. Technologically, which problems have you encountered and how you have solved them, and any processes followed that guarantee that the solution fulfills the challenge/Theme Challenges and data provider req4uirements.

Youbiquo follows international standards like ISO 9001 to ensure customer satisfaction and product/service quality improvement. We use Agile methodologies like SCRUM, conducting sprints lasting over two weeks with continuous sprint reviews. Risks in the QIMaL development concern: DATA QUALITY AND AVAILABILITY RISK: Ensuring data quality and availability is crucial in ML. Incomplete, inaccurate, or biased training data can harm model performance. This risk was mitigated through thorough data cleansing and preprocessing. MODEL OVERFITTING RISK: To prevent overfitting and ensure model generalization (e.g., CNN, CED, RNN), we used L1/L2 regularization, Dropout, EarlyStopping, Data Augmentation, and Cross-validation. OPTIMIZATION OF HYPERPARAMETERS RISK: Setting hyperparameters like batch size and learning rate was necessary for model training. We utilized algorithms like Bayesian Optimization and packages like scikit-optimize for selection. DATA STORAGE AND MANAGEMENT RISK: Handling large quantities of diverse data required scalability. MongoDB and MinIO were chosen for data and object management.

This project has received funding from the European Union's H2020 research and innovation programme under Grant Agreement no 951981
BIG DATA VALUE This project is part of BDV PPP
CORE PARTNERS
DATA PROVIDERS