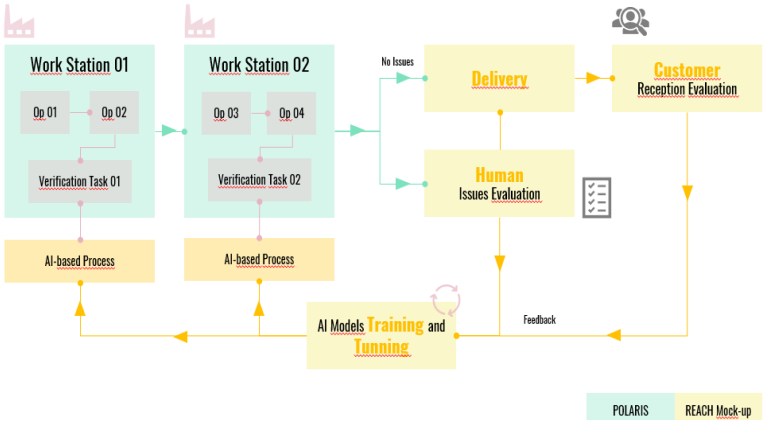
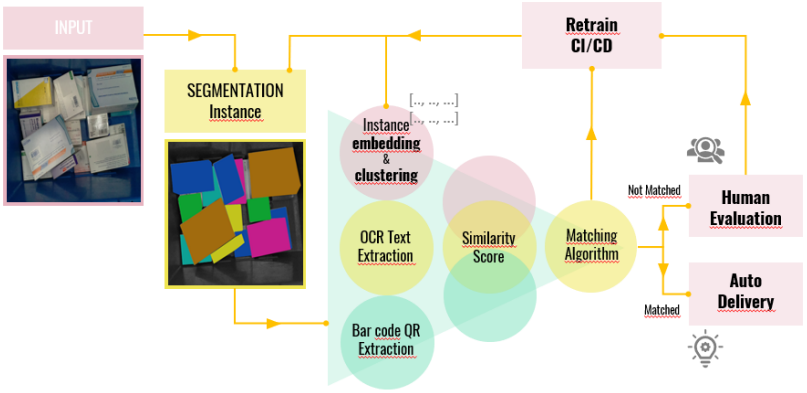
Technical Specification Double-side Page

1. **TECHNICAL SCOPE:** Summarize the mock-up devised during the EXPLORE phase: how have you addressed the challenge/Theme Challenges and tackled with its requirements and data. Include a diagram.

Polaris manages a workorder as a sequence of operations, some of them need to be verified when it's completed. That's the place in which an multitask AI-based verification process will be performed, searching for anomalies. Cofares' challenge is a particular use case. A picking order is a sequence of picking operations, for each one it's necessary to perform a verification task, searching for mismatched products respect the content of the order. At the end, if there is not any issue, an automatic delivery is executed, otherwise a human verification is needed. This human intervention, including the feedback from the final customer are used to retrain models. Green-filled processes already developed in Polaris, yellow-filled the mockups analysed in the Explore phase.





Specifically, for Cofare's challenge the AI-process is: for each picking station a picture is taken, that is the input for a segmentation to extract the object instances. Then, three parallel different tasks are used to calculate a similarity score. Image embedding with clustering for similarity, OCR text extraction and Barcode and QR code extraction. Finally, a matching algorithm is used to check the identified product with the content of the picking order. Image instances of matched products will be used to retrain the segmentation and embedding tasks.

2. **ALGORITHMS, TOOLS AND CONCLUSIONS:** Detail the algorithms and tools identified to accomplish the challenge/Theme Challenges. Show clear understanding of the used REACH dataset/s and addressed challenge/Theme Challenges.

Scalability and MLOps CI/CD. Polaris already uses K8s for scalability. We're going to introduce Kubeflow for MLOps, and using tf-jobs and tf-serving objects.

Image OCR. We use Google Tesseract for OCR and OpenCV for image management, edge detection and bar codes detection.

Artificial Intelligence. Mainly Tensorflow, wrapped with Keras and Scikit Learn.

Models. Modified R-CNN for segmentation. Other common CNN networks for image embedding and QR code detection.

Other big-data tools. Kafka and Pusher for streaming, Apache Beam as base framework for pipelines and scripts, Google Cloud Storage and Google DataFlow.

REACH resources and toolbox. We are interested in using Anonymizer, and k8s and GPU infrastructure.

3. **SCALABILITY AND FLEXIBILITY OF THE SOLUTION:** Discuss whether the solution can truly cope with humongous and increasing datasets and how flexible it is to adapt to other related domains

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951981

All Polaris modules are already deployed using Kubernetes deployments, allowing us to set the number of replicas of the services depending on the demand. The new modules developed in REACH will be deployed in the same way, as Polaris extra services. MLOps workflows will be implemented using Kubeflow, it will be used to implement data pipelines, auto training and auto deployment, using mainly tfjobs and tfservicing objects. IoT devices (RPI+coral accelerator, Nvidia Jetson) will be used as k8s agents to implement edge computing and AI tasks execution, mainly for GPU and TPU based tasks.

These technical points are the base to assure the scalability of our solution.

Regarding the flexibility, the new AI-based Polaris Module we are propossing is a general-purpose solution, it's a way to build, train and manage components to implement automatic AI-based verification tasks in any production process, no matter if that verification is classifying an anomaly detected in an image or if it is for detecting outliers in time-spent series or to detect strange values on forecasted series using a regression model. We are going to work on a specific use case for our data provider Cofares, but the solution is focused on how to implement any AI-based multitask verification process for any industry.

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

Polaris provides a high level of consistency in data since any production event is materialized and they are fully traceable and auditable. Additionally, no sensitive data is stored in Polaris, there is not any way to trace and find personal information.

However, data security is a must on industrial env. Aeronautical, space and military sectors (our specialities) are very strict on it. To address these conditions, Polaris can be deployed on-premise, on a private cloud or as a multitenancy system. Even in this last case, databases and stores are deployed separately. Additionally, Polaris implements a full authentication, roles, and permissions system, including user interfaces and APIs.

Additionally, Polaris Tx, is a module to register any data transaction using blockchain and IPFS to provide the trustfulness needed for some processes, ie delivery and reception. We'd like to compare the features of Polaris Tx with blockchain related tools included in the REACH toolbox, and get mentoring from CEA (LICIA) about the smart contracts we are using. These features will be very useful to ensure the data value chain between COFARES and its customers in order to include secured hash codes in the data of a specific automatic delivery, or to get hashed information too when customer returns feedback about specific issues in the reception. This is something common in Polaris too, industrial sector is usually a complex supply chain with delivery and reception processes for each link.

Another tool we'd like to include in our solution is Anonymizer, it's always necessary when we are sharing info with end users, as we're doing in our solution.

5. **QUALITY ASSURANCE AND RISK MANAGEMENT:** Describe the quality process planed for the final product. Technologically, which are the potential risks in all the phases of the project (design of the solution, development, testing, deployment…) and indicate mitigation plans to still fulfil the challenge/Theme Challenges and data provider requirements.

Specifically for AI components lifecycle, governance, and quality assurance we usually implement the standard CRISP-ML(Q). Aside basic quality aspects on an AI project like performance requirements, robustness, offline testing, etc. it is important to include, and it will be done in our solution, continuous evaluation task for the model on production. This evaluation will help to prevent model decay, bias, lacks, etc. Specifically for the Cofares' use case there are some identified risks:

1. Quality or accessibility of a good dataset from our data provider. To mitigate it: AI components of our solution are auto trained. The whole system can be implemented using poor datasets at the beginning. Also, our solution is a general-purpose one, so we can train general models (anomalies on production time series, quality issues severity classification, production time forecasting), models with less pre-training requirements.
2. Access good infrastructure to train image-based models using GPU. To mitigate it: Provide our own cloud infrastructure using Google Kubernetes. Request Deusto and other REACH partners for Kubernetes and GPU resources.
3. Access to the Cofares' IoT infrastructure to play with edge computing. To mitigate it: We already have RPIs, Coral accelerator and Jetson Nano devices.

This project has received funding from the European Union's Horizon 2020
research and innovation programme under grant agreement No 951981