TRUSTAI

TRANSPARENT, RELIABLE
& UNBIASED SMART TOOL

D9.1 – Management Report

INESC TEC

INESCTEC

September 3, 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under Grant agreement No 952060.

DOCUMENT CONTROL PAGE

DOCUMENT	**D9.1 – Management Report**
TYPE	**Report**
DISTRIBUTION LEVEL	Public
DUE DELIVERY DATE	30/06/2020
DATE OF DELIVERY	03/09/2020
VERSION	1.0
DELIVERABLE RESPONSIBLE	INESC TEC
AUTHOR (S)	Fábio Moreira (INESC TEC) Gonçalo Figueira (INESC TEC)
OFFICIAL REVIEWER/s	Gonçalo Figueira (INESC TEC)

DOCUMENT HISTORY

VERSION	AUTHORS	DATE	CONTENT AND CHANGES
0.1	Fábio Moreira	15/07/2021	First draft
0.2	Fábio Moreira	25/07/2021	Second draft considering partners contributions. Version submitted to reviewers.
1.0	Gonçalo Figueira	03/09/2021	Final revision

ACKNOWLEDGEMENTS

NAMES	PARTNER
Marc Schoenauer	INRIA
Peter Bosman, Tanja Alderliesten, Evi Sijben	CWI
Francisco Amorim, Teresa Bianchi de Aguiar	LTPlabs
Nikos Sakkas	APINTECH
Zehra Cataltepe, Ozden Gur-Ali, Eda Idil Coker	TAZI
Raul Vicente, Eduard Barbu, Elias Kiriakides, Saili Petti	Univ. of TARTU

DISCLAIMER:

The sole responsibility for the content lies with the authors. It does not necessarily reflect the opinion of the CNECT or the European Commission (EC). CNECT or the EC are not responsible for any use that may be made of the information contained therein.

Executive Summary

The purpose of Work Package 9 – Project Management is to ensure a smooth and timely execution of the all the planned tasks of TRUST-AI project.

This document corresponds to deliverable D9.1 – Management Report, which consists on a status report on the technical achievements of TRUST in the first nine months of the project. A brief description of the development of each task is provided, including documentation of procedures, screenshots, preliminary results and identified risks.

During the reporting period, comprehended between 01 October 2020 and 30 June 2021, all the planned tasks are being executed on a timely manner with no substantial deviations in terms of time and other resources.

Ten deliverables have been submitted with no substantial delays and two deliverables are still pending due to delays on the signing of required documents for data collection in WP5. New deadlines are proposed for the delayed deliverables.

At this point, it is difficult to connect the outputs of the first nine months of the project with the expected project impacts, since the consortium focused on literature reviews and requirements gathering tasks. The different partners have collaborated towards the joint design of a general artificial intelligence framework, which will support the resolution of the three use cases, and is expected to allow external entities to profit from most of the developments achieved in the TRUST-AI project.

Table of Contents

Abbreviations and Acronyms

AI	Artificial Intelligence
EC	European Commission
EU	European Union
PM	Person Month
WP	Work Package

1. Introduction

In this section, we present an overview of the work packages, objectives and problems referring to the reporting period comprehended between 01 October 2020 and 30 June 2021. Since this is the first project management report, most of the tasks that were carried out focused on structuring the execution process across the different work packages. In the technical-related work packages (WP1-WP7), most of the work developed by the project consortium consisted in gathering requirements for the tools and algorithms to be developed. In the non-technical work packages (WP8-W10), the work executed by the different consortium partners consisted in defining and preparing the strategies to communicate, disseminate and exploit the TRUST-AI project, while ensuring that the project complies with all the ethical-related principles adopted in the EU. All these tasks have been coordinated through a series of meetings and e-mail communications, allowing partners to self-organize in smaller groups in order to jointly develop tasks and ensure the fulfillment of the project deliverables on time.

1.1. Work packages and tasks for the reporting period

During the considered reporting period, corresponding to the first 9 months of the TRUST-AI project, at least one task related to all technical work packages (WP1 to WP7) has been started. According to the project timeline, all the tasks that should have been completed during this reporting period were indeed completed. Several tasks are now on an ongoing process, following the initial timeline with no deviations. Regarding the management work packages (WP8-WP10) where the tasks consist on the production of a few deliverables, 2 deliverables are delayed. We provide a summary of the status of the tasks that have already started in Table 1.

Table 1 - Work package tasks status summary (C – completed, O – Ongoing, D – Delayed)

Technical Work Package (1-7)	Tasks (Status)
WP1	T1.1 (C), T1.2 (O), T1.3 (O)
WP2	T2.1 (O), T2.2 (O)
WP3	T3.1 (C), T3.2 (O)

WP4	T4.1 (C), T4.2 (O)
WP5	T5.1 (D), T5.2 (D – waiting for approval)
WP6	T6.1 (C), T6.2 (D – new data may be needed)
WP7	T7.1 (C), T7.2 (O)
Management Work Package (8-10)	**Tasks/Deliverables (Status)**
WP8	T8.1 (C), T8.2 (O), T8.2 (O)
WP9	D9.1 (O)
WP10	D10.1 (C), D10.2 (C), D10.3 (D), D10.4 (D), D10.5 (C)

1.2. Project objectives and highlights for the reporting period

Since this is the first project management report, most of the technical tasks culminated in requirement specifications, which were materialized in the final deliverable of each corresponding task. The project management-related tasks defined the plans and guidelines to communicate, disseminate and exploit the TRUST-AI project, while ensuring that the project complies with all the ethical-related principles. These plans were materialized in the submission of the deliverables related to each topic. The carried-out tasks contributed in different ways to the attainment of the main objectives of TRUST-AI.

Objective 1 (O1): To develop the new foundational paradigm by designing TRUST – The meetings between the project consortium partners contributed to the consolidation of the fundamental ideas of TRUST-AI, fostering the development of human-centric and human-guided approaches incorporating generic machine learning, cognition, and interface models. Partners from different entities and research communities analyzed the feasibility of a generic

tool that is able to adequately adjust the trade-off between performance-explain ability.

Objective 2 (O2): To ensure that TRUST is adequate to tackle various problems in multiple sectors - The meetings between the project consortium partners were also important to gather requirements to apply the TRUST-AI paradigm to the three use cases, namely, healthcare, online retail, and energy. The leaders of work packages (WP1 to WP4) were constantly involved with the leaders of the use case work packages (WP5 to WP7). Therefore, the consortium partners were able to point out common lines along the three use cases and define use case-specific customizations to be integrated in the prospective TRUST-AI framework.

Objective 3 (O3): To create an innovation ecosystem around the new paradigm – Several handouts and documents are being produced in order to systematize the investigation and exploration of AI topics addressed in TRUST-AI. Dissemination is also being an important activity since the beginning of the project, and will expectedly grow as the first results are obtained. These elements will be the basis of the innovation ecosystem to be developed.

1.3. Problems encountered

During the first 9 months, the execution of the project ran smoothly and timely. Most of the planned tasks and deliverables have been executed or submitted in time. Only two deliverables are still pending due to delays on signing the required documents, particularly related to the Healthcare use case, which has to go through heavy bureaucratic processes. These issues and the new excepted delivery dates are detailed later in this document.

2. Work Package Objectives, Progress, and Achievements

In this section, we detail the objectives and the progress of each work package. A detailed description of the work executed in each task is provided, as well as a subsection explaining how the referred work contributes to the expected impacts of TRUST-AI.

2.1. WP1 (INESC TEC)

2.1.1. Overview

The purpose of Work Package 1 is to design and implement a framework compiling and generalizing all the functional requirements of TRUST-AI use case applications. To carry out this work package, the first task is to (1) gather functional and non-functional requirements suggested from discussions between the partners of each use case. In parallel, since the third use case needs a protocol for distributed and trusted data access, (2) the development of the distributed and trusted data protocol is also on an ongoing process. The last task is to (3) design and implement the TRUST framework itself. During this reporting period, T1.1 – Requirements gathering has been concluded with the submission of deliverable D1.1 - Framework requirements document. Task T1.2 – Design and implementation of the framework has already started. Multiple technologies have been compared in order to trade-off functionality with implementation efforts. We aim at implementing a web-based tool in a Python programming language, leverage the advantages of multiple packages such as FastAPI, Cron, and Vuepy. A toy app has already been implemented to test the referred technologies. Task T1.3 – Developing a protocol for distributed and multiple data sources has already started under the supervision of APINTECH, yet the integration of the protocol within the TRUST framework has still to be discussed.

2.1.2. Explanation of the work carried out

Task	Work progress and achievements
T1.1 – Requirements gathering	The requirements gathering consisted in multiple meetings with several partners in order to define the functional and non-functional requirements of the TRUST framework. Besides the framework requirements gathered from the meetings related to

	each use case, we involved the right partners, depending on their expertise, to leverage their experience on the different components of the framework, namely, AI algorithms, Cognitive models, and Interfaces. During the discussions, all partners showed a strong interest in several non-functional requirements to allow the framework to be responsive and intuitive, so that it can be potentially be used by external entities in the future (if possible). This task was concluded by the submission of deliverable D1.1 - Framework requirements document, which provides the main guidelines for task T1.2.
T1.2 – Design and implementation of the framework	The design and implementation of the framework started by documenting what should be the architecture of the application to be developed. The main objective was to design an architecture to seamlessly connect AI algorithms, Cognitive models, and Interfaces. Given that the application needs to be online and provide the possibility to run several algorithms and save the results/runs, we considered different options for database services and job scheduling technologies. At the moment, a toy app has already been implemented to run a simple AI algorithm. This toy app was mainly implemented using FastAPI, which provides several functionalities with minor implementation effort (e.g., secure login sessions, error handling).
T1.3 – Developing a protocol for distributed, multiple data sourcing & trusted data access	Work here addresses data modeling and data accessing aspects in the area of building energy management. In particular, it describes the specification and implementation of an API-application programming interface for managing real time data, along static data and meta-data. The approach is generic and is not compromised by building typologies (e.g., residential or office), scope (e.g. a sole building or the residential sector of a country), or data resolution (e.g. 1 min or a year). For sole buildings, we may often use the notation LA (local area) whereas for city/ country level approaches the term WA (wide area). This work is an important requirement for testing the XAI algorithms to be developed for typical energy applications such as demand forecasting and demand response, which are the two broad application lines envisaged in TRUST AI. As new

	concepts will emerge in the course of TRUST AI we will likely need to revisit and amend the model to address new requirements. The related deliverable is structured in two parts; PART I describes the content managed by the API as well as important, modeling considerations. Although the TRUST AI use cases are the driving force the goal is also to deliver a data sharing environment of a more general purpose and use. PART II is the actual API technical implementation where the user interfaces and functionality are described in detail. The deliverable and related API will be released on time end of September; it will support data-wise the building use case elaboration that will start about the same time.

2.1.3. Impact

During the first 9 months we focused on gathering the requirements for the framework. Logically, we cannot point out clear impacts at this point, since the implementation of the framework has not begun yet. However, we should say that the consortium partners collaborated and contributed to the initial design ideas and are totally aligned with the initial proposal of the framework.

2.2. WP2 (TAZI)

2.2.1. Overview

This work package will take the explanation content produced in WP3 (explaining the models produced in WP4) and investigate the best way to provide this content to humans. It will also iteratively improve the system's explainability, by providing a way for humans to provide input to the machine.

During this period, we worked on the explanation content and how to reflect this content with a useful interface.

2.2.2. Explanation of the work carried out

Task	Work progress and achievements
T2.1 – Preliminary studies on explanations from machine to human	Explored Local: LIME, Shapley, Global: Surrogate model explanations, Autoencoders. Conducted research on the explainable AI methods and tools for purposes of different stakeholders, including regulatory, end user affected by the decision, the business decision maker, machine learning engineer and domain experts. Identified need for XAI method between the local and global explanations providing regions of data with distinct relationship patterns and sensitivities. Conducted research on measuring interpretability. Suggested methods include measurement based on user performance on tasks that are based on understanding of the AI model. Interactive Regional LIME method development explores the use of autoencoders to determine the regions of data where a LIME model is applicable. Providing the user with suggestions for the boundaries of these regions along with their fidelity will allow the user to attain an explanation model that is in line with their mental model.
T2.2 – Development of user interfaces	Started implementing interfaces for searchable LIME. The user needs to be able to see explanations at different sets of data points of their choice and further explore closer points. The searchable interfaces will also allow an understanding of important features for different users as well as an understanding of proximity between data points.

2.2.3. Impact

No substantial contributions to the final project impacts are quantifiable in this preliminary phase.

2.3. WP3 (U. of TARTU)

2.3.1. Overview

The main objective of WP3 is to formalize human heuristics in producing causal explanations and apply such formalizations to simplify and prioritize local and contrastive explanations. To this end the WP is divided into 2 tasks. We have

focused on task 3.1 "Formalizing human heuristics for causal explanations", which is the only task that started (Months 7-18) during the reporting period. The main objective of this task is to quantify explanatory variables (or features in machine learning and genetic algorithms models) in terms of their saliency for different heuristics and biases used by humans in producing explanations.

2.3.2. Explanation of the work carried out

Task	Work progress and achievements
T3.1 – Formalizing human heuristics for causal explanations	Given the interdisciplinary nature of the present task (involving psychology, neuroscience and machine learning) we started by conducting a review on the literature for the discovery and production of human causal explanations. Based on such a review we have listed a dozen of known biases and heuristics that have been described to be used by humans in their production of causal explanations. Initial steps toward the quantification of some of these biases are being taken at the moment. The next steps in this direction include the formalization and quantification of all biases and the tuning of the weights of different biases to produce a saliency mask over all the features (explanatory variables). Also based on discussions with use case of WP6, we decided to explore not only the biases and heuristics used by humans but also the different combinations of features (operators) used by humans in their formation of higher-level explanatory variables. In relation to applications to use cases, an initial data analysis of data from WP7 (TRUST Instantiation in Energy) was performed. This analysis included explainable baseline models such as SHAP and LIME, as well as classical multivariate regression models. Based on this research, a paper in collaboration with WP7 is being written at the moment.

2.3.3. Impact

No substantial contributions to the final project impacts are quantifiable in this preliminary phase.

2.4. WP4 (INRIA)

2.4.1. Overview

(O1): To develop the new foundational paradigm by designing TRUST.

Our implementations aim at enabling presenting multiple alternative models to users. We expect that this will make important improvements towards enabling 'human-guided symbolic learning'.

(O2): To ensure that TRUST is adequate to tackle various problems in multiple sectors.

By enabling the presentation of multiple models to the user, the user can pick a model that is more appropriate towards the target domain. This may turn this method applicable to more diverse problems.

During the 3 first months of this WP, we have started developing techniques for searching for diverse high-quality solutions by implementing multiple methods. Thereby we aim to steer towards 'human-guided symbolic learning' as it enables to present multiple models to the user. In addition, implementing these methods required to implement some key elements for GP such as multitree representation and multi-objective search. This enables easy implementation of other algorithms using these elements.

2.4.2. Explanation of the work carried out

Task	Work progress and achievements
T4.1 – Fundamentally advancing GP-GOMEA	We have started developing techniques for searching for diverse high-quality solutions by implementing multiple methods: • First, by employing niching methods that operate in the semantic space of the solutions. Specifically, we cluster the population in semantic space each generation to represent niches. Subsequently, individuals can then only mate with individuals that reside in the same niche. The aim of this is to explicitly split the search bias of the genetic programming approach so that different types of solutions can be effectively and efficiently found within the different niches. • Second, by using a multitree representation (i.e., representing multiple solutions (in a vector) at the same

	time). We use such a representation together with multi-objective search and optimize for multiple solutions that each have the lowest possible error, but at the same time have the largest possible distance between them in semantic space. A multi-objective approach allows us to optimize for difference and quality at the same time, resulting in multiple vectors of solutions that constitute a high-quality set of trade-off vectors between these two objectives. We think this leads to better results than when improving both aspects separately (as is the case with niching), because then it is harder to prevent the algorithm from finding solutions that have either a low error or a large distance to other solutions because of the relation between distance and error. Moreover, having a set of vectors of trees that represent multiple equally good solutions in terms of error, but potentially different interpretations, we can not only learn more about the multimodality of the search space from the perspective of interpretations, we can also see if there is any inherent difference in these solutions that can only be judged by an expert. It is expected to be of additional value in practice to be able to present such multiple alternatives to experts. To establish this method, we have been working on implementing both multitree representations as well as a multi-objective version of GP-GOMEA. Both did not previously exist. We have found that implementing multi-objective GP-GOMEA is not a trivial extension of the previously published binary multi-objective GOMEA as during search particular phenomena occur that do not (often) occur in the binary domain. Specifically, it can easily happen that solutions remain in the population that are not near the non-dominated front. They may even be relatively far away as it is often easy to create solutions that have a very large error and/or a very large distance between pairs of solutions in the multitree solution vector. With the current way that MO-GOMEA attempts to improve and approach the Pareto front, these solutions will remain in the population without being improved. Moreover, they result in a key component in GOMEA to work properly no longer, which is a clustering mechanism that is needed to effectively distribute the search bias along the non-dominated front. Therefore, we are investigating which changes are required to the (binary) multi-objective GOMEA algorithm to overcome these issues.
T4.2 – Fundamentally	We started to implement the continuous version of MSGP, that didn't exist, as MSGP was developed for

advancing Memetic Semantic GP	boolean framework only. This will be the basis for all future works on MSGP, including multi-objective version of MSGP, to be then used in the interactive part of the project. During this period, some work has also been done in comparing alternative GP algorithms, including baseline GPs and grammar-based methods (such as context-free grammar – CFG-GP – and grammatical evolution – GE). These algorithms were also compared to an enumerator, which generates all the possible, non-redundant, dimensionally-aware expressions. The work has been conducted in a specific problem, the job shop scheduling, but has generated interesting insights into important research questions, such as the trade-off between performance and explainability when introducing dimensional awareness, and the appropriateness of an enumerator to generate small, explainable models. This might constitute a good benchmark (and possibly a starting point) for the algorithms to be developed in this work package.

2.4.3. Impact

No substantial contributions to the final project impacts are quantifiable in this preliminary phase.

2.5. WP5 (CWI)

2.5.1. Overview

Objective: To ensure that TRUST is adequate to tackle various problems in multiple sectors.

We have made and submitted a requirements document which listed the details around our use case and ideas on how the TRUST framework could be steered towards something applicable to our use case. Together with a data protection officer from the LUMC a data protection impact assessment (DPIA) has been done. The associated documents are currently being finalized. Further, the required documents for the medical ethical committee have been prepared and submitted. Unfortunately, approval on use of the data has not yet been given. We have done all the preparation possible to start gathering data once we get approval of the medical ethical committee.

2.5.2. Explanation of the work carried out

Task	Work progress and achievements
T5.1 – Requirement specification and data sourcing	Several meetings with the clinicians have taken place to acquire and refine the requirements specification and to discuss the data gathering. A requirements document was submitted (see deliverable 5.1). Together with a data protection officer from the LUMC a data protection impact assessment (DPIA) has been done. The associated documents are currently being finalized. Further, the required documents for the medical ethical committee have been prepared and submitted. A Castor database was set up, such that we can start gathering the data once we have received approval of the medical ethical committee. We expect to receive a decision from the medical ethical committee around the end of August 2021. Next, we can start gathering the data.
T5.2 – Development of initial XAI models, experts' validation and feedback to TRUST	We have not been able to start with this task since our data is not yet gathered. We expect to be able to start in October 2021.

2.5.3. Impact

Our use case is meant to test applicability of the TRUST-AI framework to applications in a medical setting (often having small data sets and envisioned to be used in risk-related decisions processes). In this, we aim to give feedback and thereby steer towards improving the design to be well-aligned with something that is applicable for use cases similar to ours. In addition, with the work package we aim to have a positive impact on the way a specific type of cancer (paraganglioma) is treated. So far, this does not make a change towards expected impacts.

2.5.4. Deviations

As we have not yet been able to start gathering data because of the fact that we are still waiting for approval of the medical ethical committee, we have not been able to start with task 5.2. To mitigate the delay as much as possible, we have done as much preparations as possible for the data gathering (e.g., setting up the Castor database), such that once we have approval we can start gathering

right away. However, a consequence of this delay might be that we will not be able to finish task 5.2 on time. We work on other tasks within the project to use time in a manner that is useful for the project. In particular, significantly more work has been put in task 4.1 than was anticipated at first.

We have not been able to complete task 5.1 since we are still waiting for approval of the medical ethical committee. Consequently, we cannot start with task 5.2 yet. In addition, this narrows and/or delays the input we can give for WP 1,2,3,4. Based on our experience with the use case and the interactions we have had with the clinicians thus far, we are able to give input but it might just not be the complete picture yet.

2.6. WP6 (LTPlabs)

2.6.1. Overview

Work Package 6 aims at providing an application of TRUST in the time slot pricing allocation, validating the applicability of the framework in the online retail industry sector. To this end, during the aforementioned reporting period, the work has been focused on tasks Task 6.1 – Requirement specification and data sourcing and Task 6.2 – Development of initial XAI models, practitioners' validation and feedback to TRUST.

2.6.2. Explanation of the work carried out

Task	Work progress and achievements
T6.1 – Requirement specification and data sourcing	The submission of deliverable D6.1 concluded task 6.1. As such, thus far, a document providing the foreseeable requirements of work package 6 was submitted, detailing the conjoint view of LTPlabs and INESCTEC on the approach, AI models, cognitive models and interfaces. The deliverable was also presented to and commented on by the remaining partners to ensure maximal cohesion on the matter. Moreover, a data-gathering phase ensured the availability of a dataset representing an instance of the problem tailored to the needs of the WTP model.
T6.2 – Development of initial XAI models, practitioners' validation	In task 6.2 the efforts so far covered the first step of the modelling approach: The Willingness-To-Pay model. The data analysis and data preparation

and feedback to TRUST	pipelines; initial machine learning models; and a study of their interpretability using available frameworks (such as SHAP) were concluded. Work has also started on alternative explanators that generate symbolic expressions using evolutionary approaches, namely genetic algorithms. The best performing ML model – of the gradient boosting machine family – outperformed the benchmarks made against naïve approaches. Simulations on top of the model's output showed no significant biases between the predicted and actual loads on the timeslots available within the dataset. The explanations provided by the state-of-the-art explanators, namely SHAP, generally matched the intuition of business experts.

2.6.3. Impact

The approach defined by WP6 entails a paradigm shift in AI with human-assisted validation of AI models. The requirements of Deliverable 6.1 were created with the end goal of assisting the interaction between humans and complex algorithms. In Task 6.2, the model's output, as well as initial conclusions from the explanations generated, were shared with business experts leading to adjustments in model training.

The online retail market has been growing at a fast pace, putting a strain on logistics operations. As such, accurately forecasting customer choices regarding delivery windows (the main target of the WTP model developed in Task T6.2) can significantly improve the planning of the delivery operations, assist in the definition of resource capacity and improve service levels.

The challenging nature of the retail use case, both on Tasks 6.1 and 6.2, demanded constant interaction within the consortium, creating grounds for collaboration and knowledge dissemination. The backgrounds in Industrial Engineering of INESC TEC and the backbone of analytical consulting from LTPlabs bring a practitioner's perspective on the applicability of the TRUST framework, perfectly complemented by the computer science expertise of INRIA, and neuro and behavioral science expertise of University of Tartu. TAZI's feedback and unique perspective on the work developed in the WP6 showcase the importance of having an interdisciplinary consortium tackle such a complex issue.

2.6.4. Deviations

Data sourcing is within the scope of Task 6.1. This data collection step was concluded for the foreseeable needs, namely for the Willingness-To-Pay model. However, new data requirements may arise from the subsequent steps in the approach - the cost to serve model and the prescriptive heuristic. As such, the scope of Task 6.2 may have to be broadened to accommodate for such unforeseen data collection and validation efforts.

2.7. WP7 (APINTECH)

2.7.1. Overview

The purpose of Work Package 7 is to derive requirements to TRUST-AI, by proposing state-of-the-art, eXplainable Artificial Intelligence (XAI) for demand forecasting in building energy applications. There are several such forecasting scenarios that will be discussed below and that will receive attention in WP7 activities. Overall, they fall into two broad categories; in the following, they will be referred to as the energy use cases (EUCs):

EUC-1 wide area and mid-term (1- 2 years) forecasting, suitable for cities and countries and related decision-makers.

EUC-2 local area and short term (1- 7 days) forecasting, suitable for building level applications, supervised by building managers, and engaging building users.

Elaborating, developing, and demonstrating these two energy use cases is the key objective of WP7. In addition, WP7 creates a key condition for exploitation activities. More than delivering prototypes on the two subcases, we envision a conceptualized open platform that wraps them up, and where various third-party developers could be able to openly share data and energy applications

2.7.2. Explanation of the work carried out

Task	Work progress and achievements
T7.1 – Requirement specification and data sourcing	The two defined use cases, EUC-1 and EUC-2, have been in detail discussed along a similar template that includes the following: BUSINESS CONTEXT: The business and social value of demand forecasting PROBLEMS TO SOLVE: The key problems encountered - The challenges and objectives - The underlying decisions and their stakeholders - The

	constraints that are typically encountered and their effective management. CURRENT APPROACH: Overview of current literature and business approaches - Overview of technology used, machine learning included NEEDS: Limitations of current approaches in terms of data and models used - Data sourcing and models envisioned - The TRUST AI added value; embedding explanations in the solutions DATASETS: Structuring and sourcing data
T7.2 – Development of initial XAI models, experts' validation and feedback to TRUST	First trials, models (linear regression, neural networks) and counterfactual analysis have been tested for the case of residential consumption. A publication by APINTECH and TARTU has been submitted. Work is ongoing and will soon shift to the short term/ building level use case.

2.7.3. Impact

There are several innovations introduced in the use case. Innovation is here considered in a 'value generation' perspective and not solely as an 'invention' related one. Thus, the relevance to impact is much more straightforward.

The first strategic impact sought is the close link between forecasting and decision support. This is the key impact from interpretability considerations, be they global or local; even features have been selected with decision support in mind. The idea and importance of actionable features has been highlighted and has practically resulted in a major new modeling direction suggested for both scopes targeted (building/ nation). We now anticipate that global surrogate models, as well as local counter-factual functionality, will make the best use of these provisions and result in a highly useful decision framework.

The second strategic impact sought is to deliver an XAI-enabled collaborative platform for energy applications. Indeed, we do not look forward to developing just for the use cases highlighted here. It is important to allow any third party will be able to upload and share functionality along with well-defined rules and access rights. These provisions drastically leverage the usability of the technology.

The third strategic impact relates also to work planned in T1.2 that seeks to develop an open API for uploading data to our collaborative platform, as discussed in the above strategic impact line. Data in energy applications are typically real-time data, collected in more and more increasing resolutions. To source the data and present them to the collaborative platform is no trivial task and managing this by file uploads is not anymore, a promising way. We also believe that this data layer is not best exploited if it is restricted to serve strictly

as a data sourcing tier alone. At a limited additional effort, it can very well emerge into an open data-sharing platform, whereby users can share energy real-time data over well-defined policies.

2.8. WP8 (INESC TEC)

2.8.1. Overview

The objective is to foster the project dissemination in order to consolidate TRUST and spread the human-guided empiricism paradigm. In addition, the work package aims to promote communication between partners and stakeholders and define future exploitation routes for the achieved results.

To attain this objective, a communication and dissemination plan is developed. This plan includes guidelines for a comprehensive project communication, somehow similar to branding guidelines, which will help to disseminate project progress and results in a brand-like manner. This will then help us to sustain project's success after the project completion.

Later in this document, we also provide a list of the main communications released so far.

2.8.2. Explanation of the work carried out

Task	Work progress and achievements
T8.1 – Communication & Dissemination Plan (CDP) and Data Management Plan (DMP)	The first version of the CDP and DMP have been submitted. In case any update is necessary to these documents, INESC TEC will update the information provided.
T8.2 – Communication, Dissemination activities and Ecosystem generation	Press releases in Turkey and Cyprus have been issued. Additionally, project partners have been invited to disseminate the information from their social media channels such as LinkedIn. A podcast plan development is initiated; it is planned to release a podcast of 15 mins average periodically to inform the community on explainable AI, as well as project developments. With this we plan to form an educated community till the project completion.

2.8.3. Impact

As the project use cases are getting prepared, we will begin to disseminate the activities. Finally, we will try to disseminate the real-life impact of our activities on a use-case basis. In that regard, we will use the communication and dissemination plan, which will help the project to be more accurate and comprehensive. Additionally, the CDP and deliverables we will create a broad awareness around TRUST-AI and foster the engagement of a community in this topic.

2.9. WP9 (INESC TEC)

2.9.1. Overview

Work Package 9 aims at coordinating the scientific and technical activities and deal with the overall administration of the project. The activities carried out in the scope of this work package are to ensure close collaboration between all the partners of the project. Multiple meetings have been scheduled to foster collaboration between partners during the development of the tasks carried out during the current reporting period. The objective was to provide a friendly environment so that everyone could know each other and understand the how the expertise of each partner could be helpful. The writing processes of the documents related to each deliverable were also coordinated through meetings and e-mails. Dates for first drafts and reviewing were defined, ensuring the participation of several project partners. The project management activities for this reporting period are finished with the submission of the management report and the financial statements.

2.9.2. Explanation of the work carried out

Task	Work progress and achievements
T9.1 – Project coordination/Scientific and Technical Management	Management report and the financial statements referring to the first 9 months is done.
T9.2 – Administrative, financial and contractual management	All the administrative, financial and contractual management have been addressed.

T9.3 – Interaction with the European Commission	Although this is the first management report, some interactions have taken place for management visibility and other issues.

2.9.3. Impact

We cannot point out clear impacts at his point, since the major implementations have not yet begun.

2.9.4. Updates to the data management plan

No updates are necessary at this point.

2.9.5. Follow-up of recommendations and comments from previous review

This is the first report.

2.10. WP10 (INESC TEC)

2.10.1. Overview

The aim of Work Package 10 is to compile ethical issues and discussion, including the debate of sociotechnical and organizational issues of Artificial Intelligence (AI) systems in general and directly related to the TRUST project. The debate conducted here will be disseminated to multiple communities, in white papers, podcasts and videos. This will guide a 'responsible research and innovation' approach in the other WPs. During the current reporting period, five deliverables were to be submitted. Three deliverables were already submitted and two are pending due to delays on obtaining authorizations (particularly from the partners from the hospital involved in UC1). Therefore, the current state of the deliverables is the following:

- D10.1 - POPD - Requirement No. 1 (Submitted)

- D10.2 - NEC - Requirement No. 2 (Submitted)

- D10.3 - H - Requirement No. 3 (Pending)

- D10.4 - POPD - Requirement No. 4 (Pending)

- D10.5 - GEN - Requirement No. 5 (Submitted)

2.10.2. Explanation of the work carried out

This work package consists on the production of 5 deliverables. There are no tasks associated with this WP.

2.10.3. Impact

During the first 9 months we focused on gathering the requirements for the various applications to be implemented during the TRUST-AI project. In terms of ethical aspects, we cannot point out clear impacts at this point, since the implementation of these applications has not begun yet. Nonetheless, given all the discussions and research executed on the ethics in AI topic, the whole consortium is now more aware and informed the ethical aspects to consider in an AI project. We believe that the EU still has a long path to go through, but spreading awareness around the ethics topic in large projects, such as TRUST-AI, is the beginning of a shift in mentality, which necessary to create a society prepared for AI-based approaches.

3. Project Progress Report

In this section we detail the deliverables, milestones, meetings and the dissemination events that took place during this reporting period.

3.1. Deliverables and milestones

There were 12 planned deliverables for this reporting period. Only 2 deliverables are pending due to delays on the signing of necessary documents. The delayed deliverables are D10.3 and D10.4, which belong to the WP10 – Ethics. Table 2 presents a summary on the status of each deliverable involved in the current reporting period.

Table 2 - Deliverables status

Deliverable	Status – Submission date
D1.1 Framework Requirements	Submitted - 30/06/2021
D5.1 UC1 Requirements	Submitted - 31/03/2021
D6.1 UC2 Requirements	Submitted - 31/03/2021
D7.1 UC3 Requirements	Submitted - 31/03/2021
D8.1 - Communication & Dissemination Plan	Submitted - 31/12/2020
D8.2 Data Management Plan	Submitted - 31/03/2021
D8.3 Project Website	Submitted - 31/03/2021
D10.1 - POPD - Requirement No. 1	Submitted - 31/01/2021
D10.2 - NEC - Requirements No. 2	Submitted - 31/03/2021

D10.3 - H - Requirement No. 3	Delayed, to submit expectedly at the end of September
D10.4 - POPD - Requirement No. 4	Delayed, to submit expectedly at the end of September
D10.5 - GEN - Requirement No. 5	Submitted - 31/10/2020

3.2. Project meetings

Due to the pandemic situation through which the project went through in the current reporting period, online project meetings have been the main collaboration mechanism. Table 3 provides a summary on the most important meetings in which several partners participated.

Table 3 – Meetings performed during the reporting period

Meeting date (leader)	Meeting purpose
15/09/2020 (INESC TEC)	Kickoff preparation
15/10/2020 (INESC TEC)	Kickoff meeting
13/11/2020 (INESC TEC)	TRUST-AI framework requirements
08/01/2021 (INESC TEC)	AI Platforms benchmark
05/02/2021 (CWI)	UC1 Requirements
09/02/2021 (LTPlabs)	UC2 Requirements
10/02/2021 (APINTECH)	UC3 Requirements
04/03/2021 (INESC TEC)	UC Requirements Wrap-up
09/04/2021 (LTPlabs)	Trust-AI Website

13/05/2021 (APINTECH)	Energy sub cases (country/ building)
26/05/2021 (INESCTEC)	Steering meeting with parallel discussions
26/06/2021 (APINTECH)	Counterfactuals

3.2.1. Project kick-off meeting

The kick-off meeting took place on the 15/10/202. In this meeting, the leaders of each work packages briefly presented their institutions, the workplan of their work package, the deliverables, and the management procedures to ensure the smooth execution of the project. Furthermore, the legal and financial department of INESC TEC presented some information on good practices and tools to facilitate the reporting process, leveraging their experience on working with EU projects.

3.2.2. WP meetings

In each work package several meetings were done. During this reporting period, the main purpose of these meetings was to reflect and define the requirements for the TRAST-AI framework, define the requirements for each use case, and to coordinate the collaboration on report and deliverables writing process.

The most intensive meetings were the ones to define the requirements of the use cases related to work packages WP5, WP6, and WP7. These meetings always required the participation of several partners, and there was always the preoccupation in having participants from the three technical pillars of the TRUST-AI framework (AI models, cognition models, and interfaces) to support the development of use case.

3.2.3. Conferences, workshops, demonstration, and other events

Since the first months of the project focused on requirements gathering, the consortium did not have enough scientific breakthroughs to share in a conference or workshop. Nonetheless, several communications have been already released as press releases, social network posts, and a webinar on explainable AI. Table 4 summarizes the main communications released during the reporting period.

Table 4 – Main communication events

Communication date	Communication description
28/08/2020 (INRIA)	News piece in the media - TRUST-AI, un projet pour une intelligence artificielle de confiance (https://ins2i.cnrs.fr/fr/cnrsinfo/trust-ai-un-projet-pour-une-intelligence-artificielle-de-confiance)
25/09/2020 (INRIA / INESC TEC)	News piece in the media - El proyecto que utiliza la teoría de la evolución de Darwin para explicar la Inteligencia Artificial (https://elpais.com/tecnologia/2020-09-25/el-proyecto-que-utiliza-la-teoria-de-la-evolucion-de-darwin-para-explicar-la-inteligencia-artificial.html)
22/09/2020 (INESC TEC)	Participation in event – Future Tech Week (http://futuretechweek.fetfx.eu)
10/02/2021 (TAZI)	News piece in the media - TRUST-AI improves trustworthiness of artificial intelligence (https://www.tazi.ai/news/trust-ai-improves-trustworthiness-of-artificial-intelligence/)
19/02/2021 (TAZI)	LinkedIn - L'entreprise turque d' Intelligence Artificielle TAZI AI Systems participera au projet TRUST AI (https://www.linkedin.com/posts/business-france-turkey_domestic-ai-receives-funding-from-eu-activity-6772123377931956224-R4WJ/)
08/03/2021 (INESC TEC)	News piece in the media - Humanos e inteligência artificial em projeto para criar ferramenta compreensível (https://observador.pt/2021/03/08/humanos-e-inteligencia-artificial-em-projeto-para-criar-ferramenta-compreensivel/)
30/06/2021 (INESC TEC)	Webinar - How should artificial intelligence help operations management in the near future? (https://www.iseg.ulisboa.pt/pt/event/webinar-how-should-artificial-intelligence-help-operations-management-in-the-near-future-2/)

4. Project Management

This section presents a summary on the workplan consolidated execution deviations. We detail the resources that were used in the first 9 months of the project and analyses the substantial deviations in terms of task execution deadlines.

4.1. Consolidated execution

4.1.1. Human resources

Personnel effort has been executed according to the initial workplan. No substantial deviations were found. Table 5 summarizes the personnel effort execution of the first reporting period.

Table 5 - Personnel effort and cost execution (PM)

WP / Partner	PM	PM Cost (€)
WP1		
01 INESC TEC	5.30	22233.01
03 INRIA	1.40	10885.97
04 NWO-I	1.14	14035.55
05 POLIS21	4.63	26286.63
07 TAZI AI	4.86	26730.00
WP1 Total	**17.33**	**100171.16**
WP2		
04 NWO-I	0.24	2960.23
07 TAZI AI	4.17	22935.00
WP2 Total	**4.41**	**25895.23**
WP3		
02 UNIV TARTU	18.74	56453.75
WP3 Total	**18.74**	**56453.75**
WP4		
01 INESC TEC	0.90	2173.00
03 INRIA	4.14	30700.93
04 NWO-I	1.17	5454.58
WP4 Total	**6.21**	**38328.51**
WP5		
04 NWO-I	4.30	18217.13
07 TAZI AI	2.25	12375.00
WP5 Total	**6.55**	**30592.13**
WP6		
01 INESC TEC	2.83	13336.00
06 Ltplabs	4.16	22880.00
WP6 Total	**6.99**	**36216.00**
WP7		
03 INRIA	0.05	668.04
05 POLIS21	4.28	21699.94
WP7 Total	**4.33**	**22367.98**
WP8		
01 INESC TEC	1.20	7975.00
03 INRIA	0.20	2560.82
05 POLIS21	1.23	5251.00
06 Ltplabs	2.20	12100.00
07 TAZI AI	1.20	6600.00
WP8 Total	**6.03**	**34486.82**

WP9		
01 INESC TEC	3.92	21775.00
03 INRIA	0.15	1878.82
04 NWO-I	0.01	169.16
05 POLIS21	0.17	726.00
06 Ltplabs	0.14	770.00
07 TAZI AI	0.19	1045.00
WP9 Total	**4.58**	**26363.98**
Total	***75.17***	***370875.56***

4.1.2. Use of resources

Project costs execution had no prejudicial deviations during the first 9 months of the project. In fact, the pandemic context of 2021 contributed to a cost reduction in terms of Travel & Subsistence costs. Moreover, hiring the necessary workforce to execute the project has not taken place yet, thus costs with equipment were relatively low for most partners.

Table 6 – Project costs execution

	Travel & Subsistence Cost Acc	Other goods and Services Cost Acc	Equipment Cost Acc	Subcontracts Cost Acc
WP1				
01 INESC TEC		5852.00		
WP3				
02 UNIV TARTU			11530.04	
WP5				
04 NWO-I	68.22			
07 TAZI AI				4439.54
WP7				
05 POLIS21		807.13		
WP8				
01 INESC TEC		36.60		
05 POLIS21		350.00		
WP9				
01 INESC TEC			35.72	
07 TAZI AI				5826.95
Total	**68.22 €**	**7045.73 €**	**11565.76 €**	**10266.49 €**
				28877.98 €

4.2. Deviations from the workplan and their impact in the project

Although two deliverables of WP10 are currently delayed, no substantial deviations from the workplan occurred. The delayed deliverables are related to necessary authorizations to use data collected from the industry partners involved in UC1 (Healthcare). However, the workplan could proceed without the need for real data, meaning that the project was not depending on this issue. We do not expect to have negative impacts caused by the delays in the authorization documents. New dates were already pointed out to conclude the two deliverables that are currently pending.

It is important to mention that the new deadlines for the two mentioned deliverables provide a good slack without hurting the project execution. Still, this situation is going to be followed closely, as the execution of the tasks related to the first XAI models is dependent on these ethics-related deliverables.