# Project Progress Report Annex

Version 20, December 2022

Foreword

# 2022-H2 Project Progress report

OMD

OPTIMAL MANAGEMENT OF DEMAND

Edited by: Hatice Betül HERSEK, PMP

Date: 24.02.2023

## Project key data

### ACRONYM and full-length title

|  |  |
| --- | --- |
| 20003 | OMD |
| Program Call | ITEA 3 Call 7 |
| Full-length Title | Optimal Management of Demand |
| Roadmap Challenge | Smart industry |

### Project duration and size

|  |  |  |
| --- | --- | --- |
| Size | Effort: 72.07 PY | Costs: 2.4 M€ |
| Time frame | Start: 2022-01-01 | End: 2024-12-31 (36 months) |

### Coordinator

|  |  |
| --- | --- |
| Türkiye | Experteam |
| Type | Small and Medium sized Enterprise |
| Contact Person | Dr. Demet Seyhan |
| Email Address | demet.seyhan@experteam.com.tr |

### Project Status

|  |  |
| --- | --- |
| Latest FPP | Change Request (21-12-2022) |
| Latest PPR | 2022 Semester 1 |
| Latest Review |  |
| Upcoming Review |  |
| PCA status | PCA has not been signed yet |

### Consortium

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Country | Funding Status | National Coordinator (Company) | Total Effort (PY) | List of Partners |
| Austria | Not funded (N) | George Suciu (BEIA GmbH) | 4 PY | BEIA GmbH |
| Hungary | Funded (Y) | Tibor Bakota (FrontEndART Software Ltd.) | 5 PY | FrontEndART Software Ltd., University of Szeged |
| Portugal | Funded (Y) | João Ribeiro Carneiro (Instituto Superior de Engenharia do Porto (ISEP)) | 19 PY | FTP - Com. Equip. Inf. Lda, Instituto Superior de Engenharia do Porto (ISEP) |
| Slovenia | Funded (Y) | Simona Brezar (Caretronic d.o.o.) | 11 PY | Caretronic d.o.o. |
| Spain | Funded (Y) | Alberto Oliva (Strategy Big Data) | 13 PY | Strategy Big Data |
| Türkiye | Funded (Y) | Arda Ödemis (ARD GROUP) | 20 PY | ARD GROUP, DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ, Experteam, Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI. |

## Project Acronyms

|  |  |
| --- | --- |
| OMD | Optimal Management of Demand |
| ML | Machine Learning |
| ITSM | Information Technology Service Management |
| NLP | Neuro-Linguistic Programming |
| GDPR | General Data Protection Regulation |
| KPI | Key Performance Index |
| DL | Deep Learning |
| AI | Artificial Intelligence |
| SDM | Service Demand Management |
| HLA | High Level Architecture |
| PCA | Project Cooperation Agreement |
| ARR | Automated Request Routing |
| SoTA | State of The Art |
| STG | ITEA Steering Group |
| min.s | minutes |
| PPR | Project Progress Report |

## Table of contents

[Project key data 4](#_Toc41573343)

[Project Acronyms 5](#_Toc41573344)

[Table of contents 6](#_Toc41573345)

[1. Project one-page description 7](#_Toc41573346)

[2. Project overall status 8](#_Toc41573347)

[2.1. Top 4 overall targeted innovations 8](#_Toc41573348)

[2.2. Top 4 overall targeted business impacts 9](#_Toc41573349)

[2.3. Top 4 overall project KPIs 10](#_Toc41573350)

[2.4. Top 4 overall risks 11](#_Toc41573351)

[2.5. Change in the technology and market during the reporting period 12](#_Toc41573352)

[3. Market access & Exploitation 13](#_Toc41573353)

[3.1. Partners’ market access 13](#_Toc41573354)

[3.2. Top 8 overall partners’ achievements 13](#_Toc41573355)

[3.3. Realised achievements 13](#_Toc41573356)

[4. Project progress during the reporting period 14](#_Toc41573357)

[4.1. Project progress and issues during the reporting period 14](#_Toc41573358)

[4.2. Details of progress per Work Package 17](#_Toc41573359)

[4.3. Per partner progress during the reporting period 17](#_Toc41573360)

[5. Additional feedback to previous STG remarks (optional) 19](#_Toc41573361)

1. Project one-page description

OMD is a product that helps businesses to assign the correct agent to a specific service demand effectively, and remotely. In our speedy internet era, it meets the need of a fast processing of each call on the demand side, and the effective management of resources on the supply side of businesses. Increasing demands and time pressures accelerated by the pandemics make organizations ask for new automations to proactively manage their environments. SDM tools are important to do so, and existing products focus mainly on IT support: ITSM. The ITSM tool market can be considered mature as in the number of products, yet in AI capabilities they are in their infancy. Meanwhile, we need advanced approaches for optimizing demand management and a better utilization of resources in many domains. However, there is no general framework providing SDM in multiple sectors such as judicial, health, sales, marketing or manufacture. The OMD will address the high demand for online service support for numerous different sectors due to the pandemics and thus will create a significant business impact.
  
OMD embodies a number of technological innovations aimed at providing cross-domain enhanced tools, components and services for efficient service demand management and remote customer support. By applying novel approaches mainly from AI, ML/DL, and NLP, OMD will significantly impact the market, providing cross-domain breakthroughs that will be validated in nine domains: software support, justice, healthcare, consumer electronics, e-commerce, telemarketing, manufactury, logistics and software development.
  
   
Our innovation applies novel AI solutions to support a general SDM framework, considering many parameters related to request, service experts (agents), customer, and company from various domains. Bringing together technology providers and use case owners from different sectors, OMD goes far beyond the state of the art. OMD strengthens the concept of Cross-Domain Cognitive Service Management for enhanced customer satisfaction, user experience, and cost savings. OMD will analyse different approaches to create efficient workflows for dynamic priority management in customer support teams. The profiling of customers and agents based on data-driven social mood analysis, will help to process new dimensions of customers, designing a methodology that captures emotions that will increase the quality of the customer experience. OMD will perform research and development in key topics: category classification, emotion classification, semantic capabilities to easily extract information from unstructured data, topic detection, demand and service level classification, intent classification, entity recognition and linking, request summarization and standardization, agent classification, solution classification and dynamic “time to finish” prediction using state-of-the-art AI and ML/DL models. Furthermore, we plan to open source some core components of the project to facilitate its cross-domain sustainability.
  
   
The OMD framework produced in this project will rapidly contribute to many sectors effectively using AI models to improve service as a CSM approach. With the remote working model now more intensely in our lives, companies providing remote support will dramatically increase in number and so our product will be high in demand.

1. Project overall status
   1. Top 4 overall targeted innovations

|  |
| --- |
| **1. Framework for a cross-domain service management software >** |

***Main contributors***: Doğuş, Hiperlink, Experteam and All

***Short description of innovation and the State-of-the-Art***:

Our innovation applies novel AI solutions to support a general SDM framework that can serve in many different domains, considering many parameters related to request, service experts (agents), customers, and companies. This kind of framework does not exist to the best of our knowledge and thus this is the main innovation from this aspect. Bringing together technology providers and use case owners from different sectors, OMD goes far beyond the state of the art.

Currently, the state-of-the-art systems focus on specific domains; ITSM being the leading domain in terms of the maturity and utilisation of the new technologies such as AI. However, even in this leading domain, the use of AI technologies is still in their infancy. OMD aims for much more effective and deep use of AI technologies in ITSM domain as well as other domains by sharing and transferring AI models among domains as much as possible. That is our innovative vision.

Within the general concept of ITSM, we focus specifically on the development of SDM platforms. Three points stand out especially in our project; ticket prioritisation, ticket integration and automated ticket routing. The successes to be achieved at these points will produce results that will positively affect the entire system.

In addition to these, one of the most powerful structures of OMD is that it has partners working with many use case models. OMD will need a cross-domain service in order to realize the interoperability model and use the OMD infrastructure for more than one use case. In this, each use case owner will need some standardization under the umbrella of OMD framework. The innovation here is to ensure that different domains with similar model structure can work together and to establish a cross domain service structure that provides project development flexibility for different domains.

|  |
| --- |
| **2. Optimization based on the dynamic workload of the agents and other constraints** |

***Main contributors***: Hiperlink and all

***Short description of innovation and the State-of-the-Art***:

Our first task is to understand the distribution of requests for a "call center" or "help desk". We will search for such data to understand how the help desk requests (tickets) are opened/sent over some time period. With this information we will be able to understand its distribution (as well as its statistical attributes: mean, standard deviation, skewness, quartiles ... .etc). This will be the first step to understand what "kind" of the data that we will be dealing with.

Determining agent workload to ensure customer satisfaction and meeting their needs while minimizing service costs is a continuous challenge. OMD will enhance prediction-based approaches, by combining not only traditional demand-supply schemas, but also different dimensions and constraints of every company. Besides agent and department workload, the profiling of agents based on emotional fingerprint (what is the mood of an agent ) and the analysis of external sources (i.e. weather ) will increase the efficiency and will allow us to provide much more detailed SLA forecasts. These innovations will be based on NLP, ML and DL approaches that will be analysed and tested during the project.

Workforce scheduling and dynamic task assignment is an optimization problem which focuses on assigning the tasks to appropriate agents for minimizing the "total" duration to complete "all / or many" tasks in a shorter time. There have been many studies on task optimization which mostly

focus on integer programming. We focus on a “mixed integer programming”1 approach as it fits our project which the method is often used to solve scheduling problems for multi-skilled agents.

Statistical metrics such that; task durations (how long a task normally takes in average and what is its standard deviation) as well as agents average resolving durations (how long does it take a ticket in category A to be resolved by an agent X, what is the average and what is the standard deviation), ( and finally, what is the std of the std of the agents) can be used as the inputs of optimization algorithm. Additionally deep learning methods such as LSTM can be used to forecast demands (number of tickets in a time frame: for example hours) which can also be used as the inputs of optimization algorithms. With this approach we aim to assign the tasks not only with the information of incoming tickets at that moment, but also with the prediction of future incoming tickets.

|  |
| --- |
| **3. Use of NLP for topic classification** |

***Main contributors***: DT, XT, USZ

***Short description of innovation and the State-of-the-Art***:

The current state-of-the-art in probably most NLP applications is to utilise a large pre-trained language model and either fine-tune it in a domain-specific downstream task with a number of labelled examples, or use a meta-training strategy, e.g., few-shot learning without actual model weight updates. The latter requires much less labelled data than the proper fine-tuning, however, works only with relatively large and carefully pre-trained language models.

The global OMD project involves mainly classification tasks in various domains. In the Software Development use case, USZ together with the use case owner, FEA develop an NLP-based AI solution to assign source code bugs to developers. This task can be considered as a classification over the available developers considering a complex set of features describing, among others, the code, the bug, and the developers themselves. The main idea in this, and also in other classification tasks is to develop AI-based solutions relying on pre-trained language models, namely one of the many variants of the BERT or the GPT families of language models. The particular choice of model and fine tuning strategy depend on the specific task at hand.

|  |
| --- |
| **4. Automated Request Routing (ARR)** |

***Main contributors***: All

***Short description of innovation and the State-of-the-Art***:

OMD will introduce innovations in several use cases by employing Artificial Intelligence (AI) and optimization techniques. These include category classification, emotion classification, semantic capabilities to easily extract information from unstructured data, topic detection, demand and service level classification, intent classification, entity recognition and linking, request summarization and standardisation, agent classification, solution classification and dynamic “time to finish” prediction using state-of-the-art NLP, ML and DL models. The project will newly introduce these advanced technologies to some of the domains or will advance the AI applications on some domains beyond the state of the art.

In customer support operations, ARR is focused on ticket routing. Ticket routing is the process of setting rules around how tickets are assigned to particular departments, agents, and functions. OMD embodies technology innovations related to enhanced, cross-domain ticket routing. OMD will perform research and development in key topics: category classification, emotion classification, topic detection/classification, intent classification, entity recognition and linking, request summarization and standardisation, agent classification, solution classification and dynamic “time to finish” prediction; generating services, components and tools that will be designed taking into account the requirements from the use cases. Intent classification is the automated association of text to a specific purpose or goal. Different classifiers categorise text into intents such as buy, change, generate complaint, etc. Emotion classification will rely on NLP-based techniques as sentiment analysis to build a model to predict the emotion of a text (or transcript) from a client. Specific attention will also be given to aspects of multilingualism as the OMD platform addresses a broad market within the EU and beyond. Innovative ML approaches like e.g. Transformer models such as BERT may be considered because they enable the transfer of learned models across languages. This will help to solve challenges of under-resourced languages in the EU.

By combining well-established ticket classifications techniques, with the most cutting-edge solutions for category classification and intent classification, OMD will add new layers/aspects for the enhanced analysis and processing of ARR pipelines. These new aspects will rely on the latest advances in Speech Analysis, NLP, ML and DL, and will also include computer-vision based services that will support the automation of tickets where paper-based/manual-based steps may be required. This will open OMD to more traditional, but widely consolidated processes currently available throughout logistics (i.e. public postal service) and judicial services.

ARR will be optimising the processes in all domains, increasing usability of the OMD Framework for the management of demand.

* 1. Top 4 overall targeted business impacts

|  |
| --- |
| **1. EqualityInJustice Business Impact** |

***Short description***: EqualityInJustice use case will enable enterprise licensing for Government Authorities including Ministry of Justice, General Directorate of Police and Union of Turkish Bar Associations. The EqualityInJustice Cloud Application is planned to be licensed for the 82 Bars and Union of Turkish Bar Associations and end user licensing model will be provided for the attorneys who will participate in assignments to counsel.

***Main contributors***: ARD

***Market / competitors***:

There are 160.651 attorneys registered to the union as of December 2021. End users (attorneys, law enforcement personnel and Ministry of Justice personnel) will be connecting the EqualityInJustice Cloud Application via the web application and mobile application, which will provide flexibility of usage. Furthermore, this business model will be targeted for other countries. There is no competition in the AI / NLP integrated demand management systems in the justice domain.

|  |
| --- |
| **2. Health care Business Impact** |

***Short description***

EU Market - existing network of distributors

Caretronic has a strong international network of distributors mostly in EUROPE and all around the world. Based on the Caretronic has developed the business model and road to market that includes market introduction to hospitals, nursing homes, care organisations, telecom companies, our international distributors’ network.

A picture containing graphical user interface

Description automatically generated

Map

Description automatically generated

***Main contributors***: Caretronic

***Market / competitors***:

Rauland, Ackerman, Tunstall

|  |
| --- |
| **3. Contact services, Telemarketing and Big Data Business Impact** |

***Short description***: Both the initial business model proposal and the estimates will have to be updated during the execution of the project. SBD proposes a sales model for contact process optimization services. OMD's services are also applicable to corporate clients and users of contact and telemarketing solutions in general. The following table shows the estimated sales of corporate and individual licences for OMD services.

***Main contributors***: SBD

SBD will integrate OMD's results as part of its strategy to automate and optimize sales agent selection and sales agent prediction. sales agent selection and resource forecasting for Telemarketing and Contact centers based on customer characteristics (Speech-to-text, NLP, Entity Extraction, sentiment analysis) and product characteristics (Target, audience and trends).

SBD will improve with OMD the customer experience, by analyzing external sources to better identify the mood, culture and socio-economic profile of a new customer in order to better match the customer with the most suitable agent and product campaign. This new strategy will have a positive impact that will scale beyond the customer interaction processes of contact services, enhancing security, providing feedback based on cognitive parameters (what is expressed, how it is expressed, what is perceived and how it feels...) and on artificial intelligence, natural language processing and Deep Learning / Machine Learning.

The Konecta Group has:

- An R&D&I strategy based on the optimization of all aspects that accompany contact processes.

- Data, based on the daily relationship of more than 71,000 agents internationally with more than 2 million daily contacts.

- A specific BIG DATA and Modeling company with experts in Data Science as STRATEGY BIG DATA.

- A proprietary S2T model, developed according to the problems of a Contact Center and telematic models.

- A distribution of operation centers in Spain based in Castilla y León where the human capital is distributed.

- A distribution of operation centers in Spain where human capital is distributed.

- A process of attracting resources and a local and international training model.

- The need to systematize learning with new real KPIS.

Additionally, the current situation has allowed:

- Large corporate clients assume teleworking as a reality.

- Homeshoring model is a necessity.

- The user experience in the contact is a variable with as much weight as the efficiency of the process.>

- SBD, bases OMD's marketing strategy on:

- The sale of new Language optimization services based on artificial intelligence, natural language processing, Deep Learning / Machine Learning and all the aspects that accompany the contact processes.

- A huge database, based on more than 55 million conversations per year, in Spanish with different accents and dialects.

- Modeling with Data Science experts in voice processing.

Based on the business strategy described above, and taking into account the discussions held with the members of the consortium, a preliminary analysis of the exploitation of results has been carried out. Both the initial business model proposal and the estimates will have to be updated during project execution. SBD proposes a sales model for contact process optimization services. OMD's services are also applicable to corporate clients and users of contact and telemarketing solutions in genera

***Market / competitors***:

Strategy Big Data S.L. frames OMD's results in the field of contact services, Telemarketing and Big Data.It is increasingly difficult for companies to differentiate themselves. Therefore, they must offer remarkable experiences to customers, leaving a positive memory. In this sense, the Call Center can be a support for business success.

The key factor is to have the necessary technology to offer a unique customer experience, meet their needs and expectations to create a bond.Many companies invest in the Call Centers to maintain the relationship with the customer, launch campaigns, sell products, answer questions and conduct opinion surveys.

Technology companies

In general, the service used by the sector is based on the helpdesk style. It has, in general, the function of providing solutions to technical problems that the customer may have. It has different levels of service.

Financial Institutions

Banks, insurance companies and credit card companies also use call center services to maintain customer relationships. The call center also becomes an alternative, new sales channel for offering investment and financing products. It also ends up facilitating the fulfillment of the objectives of professionals, who do not need to move from their work environment to conduct a negotiation.

Product sales and convergence

In addition to traditional product sales (telemarketing), providers (i.e. cable TV, Internet and telephone channels...) use the Call Center to offer products to customers and increase their chances of sales. It is based on customer prospecting, satisfaction surveys and product sales.

The contact center sector is undergoing a major change in the way it operates.

Until now, the massive incorporation of resources, manual call validation processes and/or the implementation of automation (IVRS) that flattened the demand curve sought to bring quality and profitability to the service.

The increase in process capabilities and the emergence of Machine Learning techniques has enabled a change in this model where Big Data and Inference processes allow a proactive exploitation of the customer relationship.

|  |
| --- |
| **4. Consumer Electronics Targeted Business Impact** |

***Short description***: As consumers learning and working from home, consumer electronics sales grown faster. Türkiye is the Largest TV & white goods producer in the European Market: $40 bn (CAGR: 6.5%) Export: $21 bn. To grow revenue in the domain, business analytics need insights about consumers feedback and intent after sales. There is a gap between sales and after sales customer problems for data flow because of complexity of processes on after sales support platforms or contacting call centers. According to the Turkish Call Centers Association, call centers industry had a value of TRY 5.1 billion and employed 91,000 people in all the call centers throughout the country in 2017, up from the 2013 figures of TRY 2.9 billion and 70,200 employees. The call center sector has set an ambitious target of having a work force of 300,000 people by 2023. OMD framework aims to prioritise data governance and collecting accurate data to unlock deeper insights and more cohesive decision making. Also to tolerate the complexity and make activities smart, using artificial intelligence is a must adoption at work and in our private lives. OMD will deliver practical actions to produce immediate business impact powered by AI. Such as shortening problem identification and also solution generation will reduce number of tickets and requests to call centers. In addition, 60% of the records coming to the Call Center are simple problems. Solving these problems through the call center creates a huge time and cost loss. There is no alternative to OMD for making customer support smart in consumer electronics. So the targeted business impact of OMD brings a total change to current solutions for all companies in the domain.

* 1. Top 4 overall project KPIs

|  |  |  |  |
| --- | --- | --- | --- |
|  | Initial value | Targeted value | Current value |
| **1.Reduction of average maintenance costs for tasks assigned using the developed method** | 5 hrs per task | 4 hrs per task | 5 hrs per task |

The value is computed by analysing the internal issue management system. Times spent on fixing bugs and closing the issues are recorded by the developers. We compute a simple average to determine the actual value of this KPI.

|  |  |  |  |
| --- | --- | --- | --- |
| **2. Decrease in ticket assignment time to the most appropriate agent** | 4-5 minutes | %10 (<4 minutes) | 4-5 minutes |

Ticket assignment process will include classification and processing of open tickets for the most appropriate agent in various domains. Training of the AI models will enable efficiency mainly in the classification phase and the target will be to provide less than 4 minute overall classification and processing phase for the assignment of the most appropriate agent.

|  |  |  |  |
| --- | --- | --- | --- |
| **3. Increase of customer satisfaction** | Baseline (Pre-OMD solutions) | %20 | Baseline (Pre-OMD solutions) |

Baseline will be defined as non-OMD platforms and the average client satisfaction rates of the existing solutions and their existing processes. A customer/patient satisfaction survey will be conducted in different domains and the target is to compare the pre-OMD and post-OMD satisfaction by minimum 20%. Shortened process times and more efficient agent assignments will be the basis of this satisfaction.

|  |  |  |  |
| --- | --- | --- | --- |
| **4. Increase in average number of tickets processed by an agent per month (agent utilisation)** | Baseline (Pre-OMD solutions) | %10 | Baseline (Pre-OMD solutions) |

The KPIs above which result in decrease in ticket filling and ticket assignment time will enable the agent utilisation. Pre-OMD agent performance will be compared with post-OMD agent performances on daily, weekly and monthly basis for the exact month for best comparison and evaluation.Copy the above template if more result KPIs need to be indicated.

* 1. Top 4 overall risks

|  |  |  |  |
| --- | --- | --- | --- |
|  | Severity | Probability | Stage |
| **1.** Cut-downs in funding for some countries | Medium | Almost certain | Monitoring & Controlling |

***Avoidance action***:

This risk is now realised. We could not avoid it, especially in Turkey.

***Back-up / Mitigation plan***:

We managed this risk by reevaluating the scope and reducing the number of deliverables, without compromising the content. Accordingly, partners do not relate to the project or relate less than expected. This risk then falls on Project management and coordination, who struggle with it. They cope with the consequences of resource cuts arising from budget cuts.

***A period in which the risk is relevant***

The risk is relevant throughout the project.

|  |  |  |  |
| --- | --- | --- | --- |
| **2.** Beia not receiving funding. | Low | Rare | Indentifying |

***Avoidance action***:

BEIA in search for funding.

***Back-up / Mitigation plan***:

We are in the process of evaluating a commercial proposal that we sent to an international client and which involves the development and implementation of a software solution that also integrates an AI module for online service remote IT support. The project has a launch date of December 2023. If we are declared the winner, this commercial project will bear part of the BEIA funding for the OMD project. We continue to identify new R&D calls with a theme similar to that of the OMD project

***A period in which the risk is relevant***

2023 - 2024

|  |  |  |  |
| --- | --- | --- | --- |
| **3.** Justice and Consumer Electronics use cases have risks related to GDPR. | High | Possible | Indentifying |

***Avoidance action***:

The data pool will either be anonymous or synthetic in order to be compliant with GDPR as well as the national regulations covering justice domain and issuance of personal data. Furthermore, pseudonymization will be evaluated for the encryption of direct identifiers for anonymization and allowance of processing quasi identifiers for optimised AI training.

In consumer electronics use case, a software was developed and delivered to the data provider company for masking sensitive data on dialog data.

***Back-up / Mitigation plan***:

Enabling a synthetic data pool and developing the solution according to non-usage of personal (attorney) data will provide compliance with GDPR as well as national regulations in Turkey. Different domains with various data pools will be evaluated against reverse identification threat and use-case owners will select the anonymization, pseudonymization or generation of synthetic data for the most efficient AI training.

***A period in which the risk is relevant***

The risk is relevant to all periods of the project. Possible updates in EU or Turkish regulations will be followed throughout the lifecycle of the project as well as the commercialization process.

|  |  |  |  |
| --- | --- | --- | --- |
| **4.** Cross-domain platform endangered - to integrate, in an efficient way, different techniques and proposals existing at the academic level to a particular field and use case | Low | Rare | Monitoring & Controlling |

***Avoidance action***:

Already avoided to some degree by organising the workshops for general architecture. The framework is now in the deliverable 3.1. submitted. Hence, we mitigated this. Therefore, the risk is reduced in this report.

The consortium aims to generate a set of flexible APIs - reference patterns - that allow easy connection and disconnection of services.

1. Redefining the deliverable for HLA, dividing it into two (V1: deadline M12 and V2: deadline: M18). Definition of general system architecture will be completed at the end of M18.

2. The deliverable related of T4.4 - Integration plan, will be completed at the end of M18. Until the 18th month of the project, the leader of this task will discuss with each partner relative to his use case.

***Back-up / Mitigation plan***:

Microservices are applications that can be developed and deployed individually by dedicated, self-organized teams, making it easier for multidisciplinary OMD teams to work on research activities in parallel.

Every partner focuses on their particular use-case to get the feature engineering as a first step; deadline M15. Use cases will use a standardised way to share data (MQTT, etc.)

***A period in which the risk is relevant***

* 1. Change in the technology and market during the reporting period

During the reporting period, OpenAI launched Chat Generative Pre-Trained Transformer, commonly called Chat GPT that is a generative language model. In general, the core function of a chatbot is to mimic human dialogues, in the case of Chat GPT besides handling dialogues he can write and debug computer programs, to write essays, such as writing poetry and composing music.

Chat GPT was trained on a massive data set, and can answer in different styles, and even different languages.

OpenAI already identified some of the Chat GPT limitations, like for instance:

writing plausible-sounding but incorrect or nonsensical answers;

being sensitive to tweaks to the input phrasing or attempting the same prompt multiple times;

being excessively verbose and overusing certain sentences.

We believe that Chat GPT will strongly impact future developments in this area.

Chatgpt is available for use as a paid service offered by the openai api. Apart from personal uses, it is possible to integrate it into other software and also chatbots.

It can be used by fine-tuning for tasks such as classification, sentiment analysis, etc. for chatbots used in the industry. Therefore, we can say that we will see it as an important part of the infrastructure in software soon.

The frontend doesn't need to be dialog-based, but frontend experiences can also start to move towards a simple dialog rather than buttons and menus.

It has been shown that he can pass academic specialisation exams, including medicine. In this case, the nature of the exams may need to be updated. Similarly, methods in other industries may also be updated.

ChatGPT can impact the market also by handling customer interactions quickly and efficiently. While it has this potential for effectiveness, it may not replace human creativity, or form meaningful connections with humans.

1. Market access & Exploitation
   1. Partners’ market access

|  |  |  |  |
| --- | --- | --- | --- |
| Strategy Big Data | sme | ESP | 13 PY |

|  |
| --- |
| Big Strategy data is currently part of ROCKETHALL Group. ROCKETHALL Group has a vast network of customers in Spain and LATAM, with presence in several countries such Colombia, Mexico and Peru.    Strategy Big data will integrate the project results to advance its portfolio of highly innovative, added-value niche services on AI, ML and NLP techniques. Strategy Big Data estimates that the new emerging opportunities will impact in the company with an expected employment growth, as a result of the project, of 8%, and an impact on annual turnover of 9%. A positive ROI is estimated to be achieved within 18 months of commercial exploitation. |

|  |  |  |  |
| --- | --- | --- | --- |
| Caretronic d.o.o. | sme | SVN | 11 PY |

|  |
| --- |
| RT is planning to communicate OMD project results on several scales:  Internal dissemination acitivities: Inside the company ERT will communicate activites of the project regulary  through e-mails and on a monthly basis at general company meetings. ERT will hold also task-specific  meetings with project team working on OMD project on weekly basis or whenever required.  On home-market (Slovenia) ERT has many customers (nursing homes, health-care facilities, care  organisations with primary and secondary end-userse as well as market-players) where we will promote the  system directly with personal visits and e-mail newsletters. ERT also organises business events for  presentation of novelties to existing and potential new customers from Slovenia where also OMD solution  will be presented. ERT will also publish OMD solution on their company's website.  On international markets (EU and outside EU – Turkey, Israel, Middle East, New Zealand, Brasil, Latin  America ...) ERT will present solution on exh |

|  |  |  |  |
| --- | --- | --- | --- |
| FTP - Com. Equip. Inf. Lda | sme | PRT | 11 PY |

|  |
| --- |
| Over the past few years, e-commerce has become a key part of selling physical, digital and service products. Thanks to the constant evolution of the Internet, its growth has been constant from year to year and the projections also point to the continuation of this increase. In 2021 it reached 5 billion dollars in sales and the estimate for 2026 is 8 billion dollars.  From the technological point of view, several tools and technologies have also evolved to make the use of e-commerce more accessible. And this accessibility applies not only to the consumer who makes his purchases, but also for those who sell, since there are several platforms and services that facilitate operationalizing the entire process. However, there are still many custom developed systems that allow for complete customization, but at the cost of greater complexity during development.  E-commerce will be the target of constant investment and technological evolution, because of its growing adoption and pratical use. |

|  |  |  |  |
| --- | --- | --- | --- |
| Instituto Superior de Engenharia do Porto (ISEP) | uni | PRT | 9 PY |

|  |
| --- |
| As a research center that works closely to a diversity of companies and associations, ISEP-GECAD will develop, promote, install, and undertake the maintenance of the results of this project. More in detail, ISEP-GECAD expects to develop new recommendation and search services and applications to help customers search products in web stores. Also, we will contribute to the project dissemination actions such as marketing leaflets, press releases and results presentation in conferences, journals, seminars or other events as a writer or a co-writer, as organizer or a participant. |

|  |  |  |  |
| --- | --- | --- | --- |
| ARD GROUP | ind | TUR | 6 PY |

|  |
| --- |
| EqualityInJustice use case will enable enterprise licencing for Government Authorities including Ministry of Justice, General Directorate of Police and Union of Turkish Bar Associations. The EqualityInJustice Cloud Application is planned to be licenced for the Unuion of Turkish Bar Associations and end user licencing model will be provided for the attorneys who will participate in assignments to counsel. There are 160.651 attorneys registered to the union as of December 2021. End users (attorneys, law enforcement personnel and Ministry of Jusitce personnel) will be connecting the EqualityInJustice Cloud Application via the web application and mobile application which will provide flexibility of usage. Furthermore this business model will be targeted for other countries. |

|  |  |  |  |
| --- | --- | --- | --- |
| Experteam | sme | TUR | 6 PY |

|  |
| --- |
| The solution to be developed in this project is primarily for our own use to reduce costs internally. It will attract also the attention of of our contacts who switch to the remote working model due to pandemics and wish to reduce their operation costs. Similar to Experteam we can name some twenty to thirty companies whom we have in our immediate surrounding / network as our primary market for such product. Oracles partner database lists some 150 companies in Turkey providing various services. They will be filtered for finding those sharing this particular need with XT. 50 companies for now seems a realistic estimate for this first, local market covering both Oracle and SAP partners. Our market in Europe can be estimated to be 250 companies and the global market 500 companies. We can expect some 100 of these companies to be in those countries of our consortium partners.  It makes sense to run local competitor analyses. In Turkey Asana or Atlassian could be the first competitors. |

|  |  |  |  |
| --- | --- | --- | --- |
| BEIA GmbH | sme | AUT | 4 PY |

|  |
| --- |
| BEIA already supplies big data and speech processing solutions to various customers in Romania (automotive, academia, car insurance, tourism, etc.) and will be able to sell the platform in a basic scenario as a nationwide SaaS service platform to public & private stakeholders in the profiling and analysis call-center business domain.  In an extended scenario, the commercialization of the services of the platform will be enlarged in the Balkan/Danube region through BEIA's sales and partners network. BEIA had several presentations of the solutions advanced within the project objectives, focusing on RPA for tenders. |

|  |  |  |  |
| --- | --- | --- | --- |
| DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ | ind | TUR | 4 PY |

|  |
| --- |
| With a dedicated team of researchers Doğuş Technology is working on NLP and machine learning technologies to develop digital assistants to solve/meet customer problems/demands. Doğuş Technology provide NLP services to Doğuş Group companies and outside customers. With the result of this project, Doğuş Technology targets its existing partner, Samsung. During the project and end of the project. Our customers and group companies that we work with for customer support with our other products (agent panel, Chatbots of support, sales, service appointmen) may be our first potential customers. In other sectors where product after-sales support is provided, for example, food, clothing, technology can be used. Multilingual projects use the translate option to access uncommon languages. First of all, local customers will be targeted specifically in Turkish. Subsequently, international sales can be targeted for customers with their own data in different languages. |

|  |  |  |  |
| --- | --- | --- | --- |
| Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI. | sme | TUR | 4 PY |

|  |
| --- |
| We will use the product that will come out as a result of the project as a priority in the products we have developed, to direct the errors we caught from the trace logs to the correct team and to show the correct error messages to the user. In addition, we will work effectively in projects such as licensing this product in departments such as universities, university libraries, and municipalities in terms of marketing and sales. We have studies on universities that provide American Turcology education abroad. Currently, 4 universities in America are using our products. We are planning marketing activities for our own products in the next year as the European Union countries and the UK. |

|  |  |  |  |
| --- | --- | --- | --- |
| FrontEndART Software Ltd. | sme | HUN | 3 PY |

|  |
| --- |
| By integrating the results of the project into FrontEndART's source code quality management system called QualityGate, all the users will benefit from the project's outcome. Furthermore, provided that QualityGate continuously analyzes a large number of open source software systems, the tool will be able to generate maintenance tasks for the community members automatically in an optimal way. All the functionalities (including the ones coming from the project) are free for open source communities. This will improve the productivity and reliability of open source software developments that take up a large amount of global market. |

|  |  |  |  |
| --- | --- | --- | --- |
| University of Szeged | uni | HUN | 2 PY |

|  |
| --- |
| USZ collaborates with many prominent industrial partners and research organizations worldwide, ensuring its leading position on the list of higher education institutions in Hungary according to the QS World University Rankings of 2020. As Healthcare and Health Sciences are essential parts of USZ, it will be possible to directly evaluate and use the relevant project output in a real working healthcare environment. The innovative solutions and knowledge output, on one hand, are great opportunities for USZ to strengthen its R&D activities, encouraging new collaborations and future projects. On the other hand, being a higher educational institute, USZ can also take part in various teaching activities, and help its students or the project output’s end users gain high quality, state-of-the art knowledge of a modern and innovative new technology. Furthermore, this knowledge introduced by developing new AI/ML and NLP methods during the project can be transferred to other projects and products |

* 1. Top 8 overall partners’ achievements

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 1 | Exploitation | New product | Cross-domain assignment tool | 2 |

|  |  |
| --- | --- |
| Summary | The tool will be able to combine data from different sources across multiple domains. It will generate assignments based on the selected domain information and platform. This tool will be particularly useful for users who work across multiple domains, as they will be able to create assignments based on data obtained from various domains. |
| Impact | The described tool can bring significant benefits to the companies, such as increased efficiency, improved customer satisfaction, competitive advantage, and potential for expansion. (quantification: 0) |
| Partners | Experteam, Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI., DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ, ARD GROUP |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 2 | Exploitation | New product | Reinforce and develop AI/ML models that can be applied to various use cases and scaled effectively for improved performance and efficiency. | 2 |

|  |  |
| --- | --- |
| Summary | The proposed approach involves the use of demand agents that consider user preferences and use reinforcement learning to assign tickets based on related goals in different use cases. |
| Impact | Develop and Improvement new models (quantification: 0) |
| Partners | Experteam, Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI., DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ, ARD GROUP |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 3 | Exploitation | New standard | Alpha version of purchase propensity models. | 1 |

|  |  |
| --- | --- |
| Summary | This application takes information about potential buyers from a database and tries to predict their propensity to make a purchase. The model is trained based on past behavior of buyers and takes in account their past interactions.  The model’s input comprises three data sources:  - Client’s information: Adress, Sex, etc.  - Client’s past purchases  - Past calls to the clients |
| Impact | Mejorar estrategia de ventas (quantification: 0) |
| Partners | Strategy Big Data |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 4 | Exploitation | New standard | Research of purchase prediction methodologies | 1 |

|  |  |
| --- | --- |
| Summary | State of the art studies and analysis of the different existing technologies associated to:  - Techniques and methodologies associated to purchase probability prediction.  - Techniques and methodologies associated to the generation of socio-cultural profile of potential customers.  - Techniques and methodologies associated with the prediction of rapport and mood. |
| Impact | Improving sales strategy, % (quantification: 5) |
| Partners | Strategy Big Data |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 5 | Dissemination | Conference | Footwear E-commerce chatbot System architecture | 2 |

|  |  |
| --- | --- |
| Summary | In this paper we describe the system Architecture that we are proposing to the Footwear E-commerce use case |
| Impact | (quantification: Confidential) |
| Partners | Instituto Superior de Engenharia do Porto (ISEP) |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 6 | Dissemination | Workshop | Architecture and Optimization problem discussed | 1 |

|  |  |
| --- | --- |
| Summary | We made workshops for architecture and the optimization problem for the whole consortium to agree on one general scheme that involves all |
| Impact | 1 optimization partner and 8 use-case owners are involved in this one, led by Doğus, Hiperlink and Experteam (quantification: 9) |
| Partners | Experteam, Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI., DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 7 | Exploitation | New service | Agent and product matching system for marketing | 2 |

|  |  |
| --- | --- |
| Summary | In OMD, SBD will provide new mechanisms to improve the profiling of demographic and geo-referenced data sets with potential customers. |
| Impact | New models will be generated to infer mood and socio-economic status of customers and agents, and to predict affinity (relationship, socio-economic status) (quantification: 0) |
| Partners | Strategy Big Data |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 8 | Exploitation | New product | Optimal Software Maintenance Task Assignment module | 3 |

|  |  |
| --- | --- |
| Summary | Based on the AI model developed for Optimal Software Maintenance Task Assignment, an OMD software module will be implemented (either as a library or a deployable subsystem). |
| Impact | The module will enable the integration of Optimal Software Maintenance Task Assignment feature into the management systems of the FrontEndART. (quantification: 100) |
| Partners | FrontEndART Software Ltd. |

* 1. Realised achievements

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Dissemination | Exploitation | Standardisation | New company | Patent | Human capital |
| Total: 8 | Total: 13 | Total: 0 | Total: 0 | Total: 0 | Total: 0 |

1. Project progress during the reporting period
   1. Project progress and issues during the reporting period
      1. Top 4 technical achievements

|  |
| --- |
| **1. General system architecture** |

The OMD is an application that can be rented as a SAAS service or sold as an application setup. The application is a shell that provides basic functions and will be able to support different sectors and applications through Plugins. The application will consist of parts provisioned as pods running in the Kubernetes environment. The communication between the pods and the connections will be provided by a service mesh application. The application and plugins will be expected to have their own databases. A general standardization is made now to provide flexibility for both the developers and the tenant of the infrastructure. Thus, each developer will be flexible in their model and application and data side, but the tenant will use the standard service infrastructure in order to use the OMD.

|  |
| --- |
| **2. Definition of the Optimization Problem** |

Workforce scheduling and dynamic task assignment is an optimization problem which focuses on assigning the tasks to appropriate agents for minimizing the "total" duration to complete "all / or many" tasks in a shorter time. There have been many studies on task optimization which mostly focus on integer programming. When we come to the ticket resolution business, the optimal solution will be to assign the problems to the right person at the right time and to solve them. When we look at the different use cases we have in order to reach the optimal solution, the system will need different features / inputs. For example, company wide ticket, agent information and their schedule information are among the data types that we can normally collect as raw data. However, we want to work on algorithms that we can use on the side of artificial intelligence and on deep learning algorithms on the side of the data that comes to us. We aim to predict ticket dynamic time to finish times with machine learning, regression algorithms, dynamic optimization to determine optimization based on the dynamic workload of the agents, SLAs and other company wide constraints. In addition, we aim to estimate the workload of the agents and the workload of the company / departments, again with the information of regression analysis and time series forecasting.

|  |
| --- |
| **3. AI for Justice use-case** |

OMD will bring a significant innovation to the existing attorney assignment process of the Ministry of Justice. Incorporation of artificial intelligence will allow the assignment of the most appropriate attorney for the matching case resulting in efficiency of overall assignment processes as well as satisfaction of victims/suspects and justice/law enforcement personnel. AI models will be trained according to the case descriptions with free format text data where case circumstances will be evaluated by natural language processing for attorneys’ expertise and location. Furthermore, the assignment results will be processed for increasing the performance of AI models.

|  |
| --- |
| **4. AI for Consumer electronics use-case** |

1. **Development of sentence similarity methods**

To reveal relations considering solution sentences between product support solution guides and customer-agent dialogue documents, sentence similarity techniques applied using BM25, FuzzyWuzzy, Fasttext and BERT. Several adjustments are made in the preprocessing. During the project, in case manual category labels on per dialogue obtained, it will be used to validate and enhance results of the final classification work.

1. **A software developed to mask sensitive data over customer-agent textual dialogues**

A software application developed to mask sensitive data consists of agent-customer textual dialogues. The application identifies phrases or sentences including personal data listed below:

• Customer name  
• Phone number  
• Lines including address  
• IMEI/CODE  
• Personel ID (national id number)  
• E-mail  
• Date

* + 1. Top 4 next technical targets

|  |
| --- |
| **1. Demand/service prioritization.** |

OMD will analyze different approaches to create efficient workflows for dynamic priority management in customer service teams. An intelligent analytics prioritization solution, based on the latest advances in NLP, Deep Learning, including sentiment analysis. SBD will carry out customer and agent profiling based on social mood based on analytical data -for the generation of an a priori profile- and its complementation by new voice analysis techniques to identify the socio-economic level and mood of the customer. SBD will research and develop different techniques to help process new customer dimensions, designing a methodology to capture emotions and improve customer-agent affinity.

|  |
| --- |
| **2. Agent / resource and product / ticket / demand matching system** |

SBD, Experteam, Dogus Technology, ARD and others use such function in their use-case. XT and FEA assign experts to calls as with Dogus, ARD is assigning lawyers, Caretronic agents and other use-cases assign resources.

In OMD, SBD will provide new mechanisms to improve the profiling of demographic and geo-referenced data sets with potential customers. New models will be generated to infer mood and socio-economic status of customers and agents, and to predict affinity (relationship, socio-economic status)

Sales agent selection and resource forecasting for Telemarketing and Contact centers will be automated and optimized based on customer characteristics.

In the specific case of Telemarketing, optimization plays an essential role. The proposed solution is focused on a set of services associated with optimization based on potential rapport analysis.

First, an agent will be profiled with respect to its rapport and potential affinity degree, based on cultural and sociolinguistic aspects, as well as its sales history.

Then, the affinity degree and profile of a potential new customer will be predicted in a cold call. Cold calls are currently based on databases with very simple information that generally only has a phone number, a zip code and may have some demographic information. Our proposal is to generate new components that allow, in a systematic way, to complement these databases with high added value information that allows to generate a synthetic profile, a priori, on the mood and cultural / economic / social profile for a given geographic or demographic target.

|  |
| --- |
| **3. Data Noise Reduction-Regularization** |

The large number of parameters affecting the problem of assigning the most suitable expert for the service to be given to the customer makes it difficult to determine the important parameters.

The large number of parameters affecting the problem of assigning the most suitable expert for the service to be given to the customer makes it difficult to determine the important parameters. With regularization methods, it will be possible to re-determine the weights of the parameters and to reduce the effect of unimportant parameters on the model. Regularization is also a method to correct the misleading accuracy level in case of overfitting of the model. Overfitting is one of the most typical problems any data scientist deals with. It's common for a machine learning model to perform well on training data, but not so well on test data or new datasets. This shows that the model cannot predict the output or target column of invisible data by passing noise to the output. Noise is data points that don't really reflect the true qualities of our data in the dataset, but are there by chance[Kotsilieris].

Editing greatly reduces the variance of the model without introducing large bias. As a result, the tuning parameter (α) used in regularization techniques limits the effect on variance and bias. As the value of (α) increases, the value of the coefficients decreases, reducing the variance. This increase in (α) is beneficial to some extent because it only reduces the variance without sacrificing any important features in the data, thus preventing overfitting. However, after a certain value is reached, the model starts to lose important properties, causing bias and Insufficient Fit. Consequently, the value (α) must be chosen carefully. It is a useful strategy to increase the accuracy of regression models [Friedrich].

Data collection and data preprocessing are the main causes of Overfitting. A dataset with an unequal distribution of features, noises, random data fluctuations and variance may have the opposite effect on model training. The model learns these random errors and fluctuations so well during training that the accuracy of the training data model becomes extremely high and the overfitting problem is encountered at this point. Updating weights and penalizing provides a simple solution to avoid overfitting. Table 1 shows the types of regularization and general approaches. In the article, experimental studies with Ridge, Lasso and Elastic Net methods on the basis of punishment type regularization are evaluated.

**Table 1**. Overview of regularization types, their general idea, and the statistical approaches that fall into the respective category

Graphical user interface, text, application, email

Description automatically generated

References

Kotsilieris, T., Anagnostopoulos, I., & Livieris, I. E. (2022). Regularization Techniques for Machine Learning and Their Applications. *Electronics*, *11*(4), 521.

Friedrich, S., Groll, A., Ickstadt, K., Kneib, T., Pauly, M., Rahnenführer, J., & Friede, T. (2022). Regularization approaches in clinical biostatistics: A review of methods and their applications. *Statistical Methods in Medical Research*, 09622802221133557.

|  |
| --- |
| **4. Demand Categorization** |

Experteam and Dogus Technology have stated to develop demand categorization as NLP-based classification algorithms. SBD doing this with audio data..

Machine learning-based classification methods that are planned to be used as a result of the academic literature review are given below:

- Support Vector Machine (SVM)

- Extreme Gradiant Boosting (XGBoost)

- Random Forest (RF)

- Growing Tree (Tree Boost – TB)

- Extreme Learning Machines (ELM)

- Deep Learning

- Fast Text

It is seen in the literature studies that the Radial Basis Function (RBF) is frequently used as a kernel function in the application of SVM to classification problems. As kernel functions, Linear, Polynomial and Sigmoid functions will be tried, as well as Radial Based Function. The parameters of the specified kernel functions and the optimum values of the Cost parameter of the SVM will be found by grid search.

Boosting algorithms are applied to increase accuracy in machine learning models. One of the important factors that makes the Decision Trees-based XGBoost algorithm powerful is that the tree structure created tends to minimize the error of the next tree from the previous tree. The important parameters that will affect the success performance of the XGBoost model to be developed are booster value, learning rate, gamma (minimum split loss), maximum depth, minimum number of leaves, lambda (parameter preventing excessive learning), alpha (parameter controlling the regulation of weights). . Optimum values of gamma parameter will be found by grid search. Optimum values of lambda and alpha parameters will be found by cross validation.

Important parameters that will affect the success performance of the models to be created with the RF method; maximum depth, minimum division, number of trees, maximum number of features, and minimum number of leaves. Optimization of the relevant parameter values will be carried out in order to increase the success performance of the models to be developed with the RF method.

The important parameters affecting the model success of the TB algorithm are the maximum number of trees in the series, the depth of the individual trees, the minimum size node to split, the row ratio for each tree, Huber's quantile cut, the impact pruning factor, the reduction factor, and the maximum node limit per tree. Optimization of the relevant parameter values will be carried out in order to increase the success performance of the models to be developed with the TB method.

The training speed of the Extreme Learning Machines (EML) method is considerably higher than other artificial neural network methods. EML is basically similar to neural networks with one hidden layer. Therefore, the working principle of EML is to a certain extent the same as the working principles of Artificial Neural Networks. However, in EML methods, the weights in the hidden layer are randomly assigned and these values do not change (update) in the next stage of the training. On the other hand, the weights between the hidden layer and the output layer are determined analytically and quickly with the help of a linear model in one go. The important parameters affecting the success performance of the EML model to be developed are the activation function and the number of neurons in the hidden layer. Apart from this, there is no need to enter any other parameter values during the training. Optimal values of these two parameters will be found in the algorithms to be developed.

Feature selection is an important factor affecting the success performance of classification algorithms. Feature selection is the process of selecting and finding the most useful features in the data set. Unnecessary attributes; It can increase the training time of the model, cause the model success to be high in the training dataset due to overfitting but low in the test dataset, and in cases where the data to be received in the test dataset are not similar to the data in the training dataset, the error rate in the model is high.

For these reasons, Feature Selection algorithms used in classification problems will be used in order to improve the performance of the models to be developed within the scope of the project. The algorithms planned to be used are given below:

- Principal Component Analysis (PCA)

- Linear Discriminant Analysis (LDA)

- Correlation Based Feature Subset Evaluation (CFSE)

In order to measure the success performance of the models, Precision, Sensitivity, Accuracy and F-score metrics, which are frequently used in the literature in classification problems, will be used.

In consumer electronics use-case, the dialog datasets are examined for the first 6 months. Later, also new data is obtained, these include the records opened by the personnel for customers applying to the service centers and the details of the operations that follow this process, the question and answer sets on the support web page, lastly the agent guides. There are not enough tagged user reviews in them.

Since we do not have labeled training data for using in textual similarity model, we will prefer bi-encoder architectures instead of cross-encoders.

We will create user data by performing data augmentation on existing data sets. Also synthetic data generation with generative language models such as GPT.

After this similarity based question answering work, we will then work on classification modelling.

In ARD use case, a demand refers to a judicial case to which an appropriate attorney is required to be assigned. In this scenario a match (between the case and an attorney pair) classification is considered. To quantify the level of appropriateness, the task is defined as a computational regression problem where a set of attributes from case descriptions and attorney profiles will be mapped into a matching score. Since the pair attributes will be a mix of binary, categorical and continuous variables, the latest state of the art methods used for similar situations will be used to implement this regression. Some of them are: Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Transformer based models (BERT, GPT-2, RoBERTa).

There are several popular libraries for text classification. Some we may be using or experimenting with include:

* NLTK (Natural Language Toolkit): (It has tokenization, stemming, tagging features. Techniques we can use with this toolkit: Naive Bayes, Decision Trees and Maximum Entropy classifiers, SVMs, k-NN).
* Scikit-learn: Provides a set of algorithms for text classification, including Naive Bayes, Support Vector Machines (SVMs), and Random Forests.
* TensorFlow: (CNNs, RNNs).
* PyTorch: (CNNs, RNNs and Transformer models).
* ML.NET (FastTree, LightGBM, Logistic Regression, SVM).

We plan to improve our method’ ensemble methods (bagging, boosting, stacking) and maybe try some transfer learning techniques. Our experiments will also include some feature engineering, text preprocessing and hyperparameter tuning techniques to improve performance in our particular case. As the first algorithm, it is planned to use the Extreme Gradient Boosting (XGBoost) algorithm to implement the regression.

Gradient boosting refers to a class of ensemble machine learning algorithms that can be used for classification or regression predictive modeling problems. Ensembles are constructed from decision tree models. Trees are added one at a time to the ensemble and fit to correct the prediction errors made by prior models. This is a type of ensemble machine learning model referred to as boosting. Models are fit using any arbitrary differentiable loss function and gradient descent optimization algorithm. This gives the technique its name, “gradient boosting,” as the loss gradient is minimized as the model is fit, much like a neural network. XGBoost is a sparsity-aware extension of gradient boosting for parallel tree learning (Chen and Guestrin, XGBoost: A Scalable Tree Boosting System, KDD'16).

* + 1. Top 4 issues

|  |
| --- |
| **1. Design assumptions** |

***Details***:

Assumptions made in the design phase of each use case are invalid

***Impact***:

High

***Mitigation action***:

Implement prototypes to run questionnaires and measure user satisfaction

|  |
| --- |
| **2. PCA** |

***Details***:

PCA has not been signed yet

***Impact***:

High

***Mitigation action***:

PCA is already validated by all partners and the signature process is starting.

|  |
| --- |
| **3. International Cooperation** |

***Details***:

Level of international cooperation

***Impact***:

High

***Mitigation action***:

Implementation of regular WPs meetings and f2f meetings of the International consortium.

|  |
| --- |
| **4. Privacy regulation** |

***Details***:

Justice use case has risks related to GDPR

***Impact***:

High

***Mitigation action***:

Enabling a synthetic data pool and developing the solution according to non-usage of personal (attorney) data will provide compliance with GDPR as well as national regulations in Turkey.

* + 1. Status of deliverables

|  |
| --- |
| **[Planned] What is the total number of deliverables in the project?** |

There are twenty-four deliverables defined in the overall project.

|  |
| --- |
| **[Planned] How many deliverables are supposed to be finalised (from the start of the project until the end of this reporting period)?** |

The following deliverables are supposed to be finalised below.

|  |  |
| --- | --- |
| D1.1 – Project Progress Report - PPR1 | Jun 2022 |
| D2.1- State-of-the-art Analysis | Jun 2022 |
| D2.2 - Scenarios and Use cases | Jun 2022 |
| D2.3 - Legal, ethical and Acceptance analysis | Sep 2022 |
| D2.4 - Requirement analysis | Sep 2022 |
| D3.1 - General system architecture | Dec 2022 |
| D1.2 - Project Progress Report - PPR2 | Dec 2022 |

|  |
| --- |
| **[Actual] How many deliverables have already been finalised (from the start of the project until the end of this reporting period)?** |

We finalised the following five deliverables.

|  |  |
| --- | --- |
| D1.1 – Project Progress Report - PPR1 | Jun 2022 |
| D2.1- State-of-the-art Analysis | Jun 2022 |
| D2.2 - Scenarios and Use cases | Jun 2022 |
| D2.3 - Legal, ethical and Acceptance analysis | Sep 2022 |
| D2.4 - Requirement analysis | Sep 2022 |
| D3.1 - General system architecture | Dec 2022 |
| D1.2 - Project Progress Report - PPR2 | Dec 2022 |

|  |
| --- |
| **[Delayed] Are there any deliverables delayed more than 2 months in this reporting period? If so, please explain why.** |

There is no delayed deliverables.

* + 1. Statement on project progress during the reporting period

OMD Project has continued in this period with global and national meetings, workshops on architecture were carried out. Our PCA document is negotiated and ready to be signed. Change request is submitted and now approved. We organized an in person meeting in Istanbul as the first general assembly. STG Review is now beeing prepared.

* 1. Details of progress per Work Package

#### WP 1: Project Management

Coordinating the consortium, communications management (National, International, Internal), scope and time management, risk management, resource management, configuration management, reporting were amongst the management efforts realized by Experteam in this second reporting period. We started with the monitoring and controlling of deliverables through our Jira. This shall serve also to connectivity, security and privacy. The task leader opens the deliverable and once we complete a deliverable in the drive collectively, the WP leader submits it to our jira.
  
We also led the optimization meetings and the general architecture meetings this semester.
  
We caught up with the schedule despite more than 3 months' delays and late starts of 60% of the partners.
  
The only delay is in the PCA which is now in negotiation.

#### WP 2: Use Case Requirements and Business Models

Work Package 2 incorporates the use case requirements and Business Models for OMD Framework. Within 2022 Semester 2, International consortium has completed the work package tasks indicated below :
  
   
OMD use cases from various domains are under the scope of GDPR (General Data Protection Regulation) and OMD international consortium addressed this issue within the studies under Task 2.3 State-of-the-Art Analysis and released D2.3 Legal, Ethical and Acceptance Analysis.
  
   
Task 2.4 Requirement Analysis studies were conducted covering all domains and their stringent requirements. Task defined the deliverable scope and objectives which continued with requirements elicitation. Key Performance Indicators were correlated with the OMD Framework and the use cases from each domains. Functional and non-functional requirements were identified and elaborated; providing a basis for the general system architecture of OMD Framework. D2.4 Requirement Analysis was completed and submitted within this Semester.

#### WP 3: System Architecture Design and AI Models Development

This WP will drive the design of the overall system architecture with detailed definition of the use cases and their requirements in WP2. The objective is the design of the OMD architecture and AI models. designing and implementing new pipelines and AI models to provide improved tools for matching agents, customers and campaigns. The optimization process will have as its main objective to predict not only the best agent-customer-campaign match, but to provide useful information - Insights - on shopper trends when managing campaigns.
  
In the first year, the general architecture of the system has been designed based on the architecture of microservices, services and AI models. In this task, the design of the different modules and services associated with the identified components has been carried out, generating an overview that allows the orchestration and coordination of the entire system.
  
- AI Geo-Demographics Profiler
  
- Mood/Cultural/Socioeconomic enhanced customer profiler
  
- AI Agent Enhanced Profiler
  
- Rapport Prediction-based Call generator
  
A set of flexible APIs -reference patterns- will be generated to allow easy connection and disconnection of services. These APIs must ensure the scalability and extensibility of the developed services, and will be shared and agreed with the rest of the consortium partners.

#### WP 4: System Development and Integration

During the reporting period we started the implementation of the overall system framework and the modules of the OMD system to meet the project requirements based on the design in WP3. The main focus was on the pre-processing activities.

#### WP 5: Demonstration, Commercialization and Standardization

With WP5 Demonstration, Commercialization and Standartization has started in Month 13 and will last until Month 36.

#### WP 6: Dissemination&Exploitation

Experteam carried out the coordination of the international consortium, organizing and leading the meetings and the communication management.
  
Experteam currently works on Data Noise Reduction through Regularization in Supervised Machine Learning Algorithms, towards a publication.
  
The website (http://www.omdproject.com/) was under construction by ARD and is completed this month.
  
Frontendart engaged Innoconsult, a consultant company, for dissemination and exploitation.
  
A press release on the ongoing OMD project and the participating Hungarian consortium partners was released in the Hungarian press.
  
"Footwear E-commerce chatbot System architecture" is prepared by ISEP as a conference proceeding.

* 1. Per partner progress during the reporting period
     1. Partners’ main contribution and effort

|  |  |  |  |
| --- | --- | --- | --- |
| Partner | Planned effort (Project start~ end of reporting period) | Actual effort (Project start ~ end of reporting period) | Contact |
| ARD GROUP | 1.9 | 1.9 | Arda Ödemis |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nAs Work Package 2 leader, ARD provided coordination among the OMD partners as well as contributing Tasks 2.3 State-of-the-art Analysis, 2.4 Requirements Analysis and provided inputs for the deliverables D2.3 Legal, ethical and Acceptance Analysis and D2.4 Requirement Analysis. As a use-case provider with EqualityInJustice, ARD contributed to T3.1 General System Architecture Design, ensuring the cross-domain operability of the solution. OMD web page has been designed and published |
|  | Discrepancy explanation:\nThere is no discrepancy between the planned effort and actual effort. |

|  |  |  |  |
| --- | --- | --- | --- |
| Experteam | 1.3 | 1.3 | Demet Seyhan |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nIn this period XT as project leader led the general meetings, both national / international; acted also as a technical leader, organized / led architecture and optimization meetings, discussed the general system architecture and the optimization problem to be defined.  As project manager, XT started monitoring progress through own jira.  As partner, data cleaning is realized in the technical team, particularly Data Noise Reduction through Regularization has started.  XT also leads the PCA now. |

|  |  |  |  |
| --- | --- | --- | --- |
| DOGUŞ BILGI ISLEM VE TEKNOLOJI HİZMETLERİ | 0.93 | 0.62 | Setenay Gemici |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nNLP model development using sentence similarity techniques applied consisting of BM25, FuzzyWuzzy, Fasttext and BERT that will be used to validate and enhance results of the final classification work. Several adjustments are made in the preprocessing during the development. A software application developed to mask sensitive data of agent-customer textual dialogues using regex and wordlists. |
|  | Discrepancy explanation:\nThe start of the project was delayed but planned research was carried out in the desired time during the period. There was a cut in the national funding proposal. (for PPR1 explaination)  Since personal data is available in the data we have, masking studies had to be completed. |

|  |  |  |  |
| --- | --- | --- | --- |
| Caretronic d.o.o. | 4 | 4 | Simona Brezar |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nWe examined a set of Slovenian cases for the use of the solution in the field of healthcare. We did the design and analysis, and in the area of ​​the software solution, we did the design and analysis of the case for the field of healthcare. Detailed description of use cases.  Solution enables healthcare professionals, nurses, logistics, maintenance issues and automatically categorizes using AI models. The solution will offer ticketing for all maintenance work (technical breakdowns, cleaning,..) |

|  |  |  |  |
| --- | --- | --- | --- |
| Instituto Superior de Engenharia do Porto (ISEP) | 2.1 | 4.95 | João Ribeiro Carneiro |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nISEP participated in all of the project’s meetings, and contributed to the project's reports and other management activities. ISEP contributed to them following deliverables from the perspective of the Portuguese use case.  D2.3 -Legal,ethics and acceptance  D2.4 - Requirement Analysis  D3.1 - General system architecture  D3.2 -Data Pipelines and pre-processing design  D3.3. AI models design  D4.1 AI model pipelines pre-processing  D4.2 AI models  D4.4. User interface |
|  | Discrepancy explanation:\nThe Portuguese Consortium is only funded until mid 2023, so there was a need to increase the effort in 2022. |

|  |  |  |  |
| --- | --- | --- | --- |
| FTP - Com. Equip. Inf. Lda | 2.61 | 5.1 | Germano Fernando Santos Pinto |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nDuring the reporting period, we participated in all meetings, reports and deliverables.  The contributions of the Portuguese use case were reproduced in the following deliverables:  D2.3 -Legal,ethics and acceptance; D2.4 - Requirement Analysis; D3.1 - General system architecture; D3.2 -Data Pipelines and pre-processing design; D3.3. AI models design;  As WP4 leader, we conducted the coordination with the consortium for the deliverables D4.1; D4.2; D4.3; D4.4. |
|  | Discrepancy explanation:\nPortugal National program is only funded until June 2023. For that reason we need to increase our effort. |

|  |  |  |  |
| --- | --- | --- | --- |
| Strategy Big Data | 2.25 | 3.68 | Alberto Oliva |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nA set of use cases and test environments have been generated for OMD. The use cases are associated to the real context of telecontacted services and will be used as a platform for pilots.  The general architecture of the system has been designed based on the architecture of microservices, services and AI models, SBD will carry out the design of the different modules and services associated to Telemarketing. SBD will generate a set of flexible APIs. |
|  | Discrepancy explanation:\nThe novelty component of the development, together with the DMO stratification factor for SBD, has meant investing more resources than initially foreseen. However, the subsidy obtained will correspond to that initially requested. |

|  |  |  |  |
| --- | --- | --- | --- |
| Hiperlink Eğitim İletişim Yayıncılık Gıda San. Paz. ve TIC. LTD. STI. | 0.75 | 0.75 | Hilmi Oğuz |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nStudies were carried out on what kind of structure the infrastructure architecture should be in multiple use cases and what kind of architecture it should be.  Studies were carried out on the optimization tool of the infrastructure. These studies were carried out on how to optimize the infrastructure, what inputs and outputs are needed, and the minimum requirements for optimization were determined and shared with the consortium.  Work was carried out for the general planning meetings. |

|  |  |  |  |
| --- | --- | --- | --- |
| FrontEndART Software Ltd. | 0.37 | 0.69 | Tibor Bakota |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nFrontEndART determined both functional and non-functional requirements against the OMD platform that are required to carry out the OSMTA use case. We also took part in reviewing partners' requirements and finalizing the whole requirement set. FrontEndART designed the architecture (in details) for the OSMTA use case, then integrated it in the global OMD architecture. We took part in the review of the global OMD architecture. Our detailed designed will be directly used for our further activities. |
|  | Discrepancy explanation:\nThe discrepancy between the planned and actual effort has two reasons. First, the experienced developer we planned to assign to the project were not available and had to be replaced by a less experienced colleague, thus the activities took more time. Second, the architecture planning on our side was deeper and well founded, and we will be able to reuse some results directly in later activities. |

|  |  |  |  |
| --- | --- | --- | --- |
| BEIA GmbH | 0.95 | 0.3 | George Suciu |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nWP1: Beia participated in the planned online meetings.  WP2: Beia continued to work on state of the art analysis. Beia worked on the definition of the assigned use-case and on the test scenarios of the OMD platform for this use-case.  WP3: BEIA started working on time-critical cloud architecture and models for timeseries from IoT/telemetry logs.  WP6: Beia continued to disseminate the OMD project at various events where Beia participated. |
|  | Discrepancy explanation:\nFunding was not obtained so BEIA is working self-funded on a reduced best-effort basis. |

|  |  |  |  |
| --- | --- | --- | --- |
| University of Szeged | 0.31 | 0.23 | László Vidács |

|  |  |
| --- | --- |
|  | Main contributions during the reporting period:\nUSZ contributes to WP3 with a document related to T3.3 and T3.4. We reviewed the state-of-the-art in NLP focusing on the Software Development domain. Latest methods, AI models, and datasets related to developer assignment tasks are reviewed. We discussed the usage of knowledge graphs in Software Development as well as methods for the developer profiling task. The document serves as the technological and theoretical basis for the Software Development use case. |
|  | Discrepancy explanation:\nDue to the late decision of the national support USZ started the tasks with a few months delay. At the same time, the actual management cost for finishing milestone 1 tasks were lower than initially planned. |

* + 1. Actual vs. planned effort overview

|  |  |  |
| --- | --- | --- |
| Report | Planned effort up to reporting period (PY) - total: 72.07 PY | Reported actual effort up to reporting period (PY) |
| 2022 Semester 2 | 17.47 (24% of total) | 23.52 |
| 2022 Semester 1 | 8.735 (12% of total) | 9.37 |

1. Additional feedback to previous STG remarks (optional)

**To STG reviewers**: This chapter is meant to provide additional information on the status of actions, in addition to the information on the online action tool (the information is exported on the Excel file). The project consortium uses this chapter to provide longer and more detailed information that are too exhaustive for online action tool and the Excel export.

The project actions for PPR1 are accomplished in the PPR2 report and the actions are solved in ITEA Portal, the actions shown as below.

Action OMD-41

Improve and extent your market and competitor assessment for the "Health Case Business" case

Caretronic 30/Dec/22

Done.

Explanation : We improved this part in the Section 2.2 of this PPR2 report.

Action OMD-40

Try to assess the initial value of KPI#4 - which could e.g., be the avg. Value over all current ticket filling times

Doğuş Technology 30/Dec/22

Done.

Explanation : Metric has been removed in the CR. Dogus Technology considers re-adopting it in the process, as the reference value will be measured within this work.

Action OMD-39

The description of Risk#2 should be improved; Risk#4 is unclear and should be improved

BEIA 30/Dec/22

Done.

Explanation: In section 2.4 of the PPR2 we improved and clarified these risks.

Action OMD-38

Please clearly state your "achievements" and especially what has been already accomplished and what is currently in the pipeline/planned

Experteam 30/Dec/22

Done.

Explanation : In this PPR2 report we clearly state our achievements in terms of what has been accomplished and our next targets.

Action OMD-37

Please improve the next 4 technical targets for next PPR and especially update mitigation plans as they are weak

Experteam 30/Dec/22

Done.

Explanation : In this PPR2 report we improve our technical targets and mitigation plans to make them stronger.

Action OMD-36

Update progress per WP, especially for WP2

ARD Grup 30/Dec/22

Done.

Explanation : Could not update directly from the ITEA portal. However, the MS Word document contains all the additions that are marked in the red color. This document is uploaded as Appendix1 in the PPR2 report.

Action OMD-35

For partner DOGUS please provide a description of your contribution in Partner's main contribution

Doğuş Technology 30/Dec/22

Done.

Explanation : Dogus contributed much to the General system architecture in both periods. Both periods' partner main contributions are now covered in PPR2, with the sum total of their yearly effort.

Action OMD-34

Maybe think on including also monetary KPIs that will allow you to better sell you solution in the future

SBD 30/Dec/22

Done.

Explanation : We do not feel comfortable defining an economic KPI. However, we have some forecasts on the number of licences we expect to sell. We added these projections to the Section 2.2. (Top 4 overall targeted business impacts) in this PPR2 report.

Action OMD-33

Think on improving this statement on project progress as e.g., there are many delays due to late start of partners

Experteam 30/Dec/22

Done.

Explanation : We have already caught up with the planned schedule of OMD. Thus, we have improved our statement on delays in this PPR2 report.

Action OMD-32

Consider updating a CR to rebase your commitments and update the workplan/milestones/deliverable due dates

Experteam 30/Dec/22

Done.

Explanation : CR is approved with the mentioned changes.

OMD-31

Please try to remove blanks and empty lines in the template as they cause misleading '<' and '>' signs in the generated PPR document.

Experteam 30/Dec/22

Done.

Explanation : We are careful about the blanks and signs in the PPR2.