**ITEA 3 Call 7**

**OMD**

**Optimal Management on Demand**

Project Number : ITEA-2021-20003

Project Start Date : 01.01.2022

Project Duration : 36 Months

**State-of-the-Art Analysis**

**References :**

Work Package : WP2 Use Case Requirement and Business Models

Task : T2.3 State of the Art Analysis

Deliverable Document : D2.1 State-of-the-art Analysis

Version : V0.2 Jun 15, 2022

Dissemination Level : Public

**Authors :** in alphabetical order

| **Company** | **Name Surname** | **E-mail** |
| --- | --- | --- |
| ARD Group | Bilge Başdemir  Esin Soycan  Özgün Sapmaztürk | bilge.basdemir@ardgrup.com.tr  esin.soycan@ardgrup.com.tr  ozgun.sapmazturk@ardgrup.com.tr |
| BEIA Gmbh | George Suciu | [george@beia.ro](mailto:george@beia.ro), [stefan.stefanescu@beia.ro](mailto:stefan.stefanescu@beia.ro), |
| Caretronic d.o.o | Simona Brezar | simona@eurotroni |
| Dogus Technology | Utku Colak  Alper Gun | [utku.colak@d-teknoloji.com.tr](mailto:utku.colak@d-teknoloji.com.tr)  [alper.gun@d-teknoloji.com.tr](mailto:alper.gun@d-teknoloji.com.tr) |
| Experteam | Taner Ulusinan | [taner.ulusinan@experteam.com.tr](mailto:taner.ulusinan@experteam.com.tr) |
| Frontendart Software | Tibor Bakota | tibor.bakota@frontendart.com |
| FTP - LDA | Germano Pinto  Isabel Ribeiro | [germano.pinto@ftpporto.com](mailto:germano.pinto@ftpporto.com)  isabel.ribeiro@ftpporto.com |
| ISEP | Goreti Marreiros  Joaquim Santos  Pedro Carrasco | [mgt@isep.ipp.pt](mailto:mgt@isep.ipp.pt)  [jpe@isep.ipp.pt](mailto:jpe@isep.ipp.pt)  rasco@isep.ipp.pt |
| Strategy Big Data | Alberto Oliva | alberto.olivia@strategybigdata.com |
| University Of Szeged | Laszlo Vidacs | lac@inf.u-szeged.hu |

Summary :

OMD is a software tool that helps service providers from various domains to use their resources effectively, to provide agile solutions, to conserve their knowledge base, to allocate each call unmistakably to the right expert / provider, and with the best solution of the particular problem at hand. This analysis includes an extensive literature and industry review on existing Technologies and the norms in order to ensure that OMD is built on the most recent technology. AI models including NLP, Deep Learning/Machine Learning and optimization will contribute to many sectors for increasing the automation level.

Table of Contents

[1. Introduction](#_heading=h.30j0zll) **4**

[2. State of the Art Analysis](#_heading=h.1fob9te) **5**

[2.1. ARD Group](#_heading=h.3znysh7) 5

[2.1.1 Market analysis for Justice use Case](#_heading=h.2et92p0) 5

[2.1.2 Literature Survey, Definition of the Problem and Solution Approach](#_heading=h.tyjcwt)

[for Justice Use Case](#_heading=h.tyjcwt) 5

[2.2. BEIA GmbH](#_heading=h.3dy6vkm) 7

[2.3. Caretronic d.o.o.](#_heading=h.1t3h5sf) 8

[2.4. Dogus Technology](#_heading=h.hph3rpf2wxnt) 8

[2.5. Experteam](#_heading=h.17dp8vu) 11

[2.5.1. Market analysis for ITSM use case](#_heading=h.cji6g5vbx7yu) 11

[2.6. FrontEndART Software](#_heading=h.3rdcrjn) 14

[2.6.1. Market analysis for Optimal software maintenance task assignment use case](#_heading=h.yf36n08ewrt3) 14

[2.6.2. State of the art for Optimal software maintenance task assignment use case](#_heading=h.sl326wt8dkgr) 14

[2.7. FTP-LDA &ISEP](#_heading=h.26in1rg) 15

[2.7.1.](#_heading=h.mlbnh16bf5ck) Market analysis for e-commerce use case 15

[2.7.2](#_heading=h.wtrrev6qzie4) State of the art for e-commerce use case 16

[2.7.2.1](#_heading=h.77hl56j6ynxj) Tools to create chatbots 17

[2.7.2.2.](#_heading=h.qm3enbbyyam4) Related works 17

[2.8. Strategy Big Data](#_heading=h.1ksv4uv) 20

[2.9. University of Szeged](#_heading=h.44sinio) 21

[Literature review and state-of-the-art in NLP](#_heading=h.1pts1twhgrfu) 21

[General introduction to NLP](#_heading=h.hrdhcekxp9ta) 21

[State-of-the-art in text embeddings and representation](#_heading=h.h5lrnu6dwe9o) 22

[3. References](#_heading=h.2jxsxqh) **25**

[**ARD Group**](#_heading=h.ci2ti7obslkj) **25**

[**FEA**](#_heading=h.y8dn4gsr6sa2) **25**

[**USZ**](#_heading=h.lbnouy4s92nt) **25**

[**SBD**](#_heading=h.dtak4qwqxer6) **26**

**Dogus 26**

**Experteam 27**

**FTP&ISEP 27**

# Introduction

The state of the Art Analysis consists of each contributor’s use case and task based research and analysis of literature and a review of industry, for ensuring OMD to be built on most recent technology. AI models including NLP, Deep Learning and Machine Learning will be selected as the key disciplines of this study in order to reach the target of increasing the automation in OMD’s use cases.

Companies’ use cases in this study will cover :

* ARD Group : Justice Sector(EqualityInJustice)
* BEIA GmbH : Logistic and Operations Support
* Dogus Technology : Consumer Electronics (SmartFix)
* Caretronic d.o.o. : Healthcare
* Experteam : Software Support (Tickota)
* FrontEndART : Software Development (Optimal software maintenance task

assignment)

* FTP-Com. Equip. Inf. Lda : E-Commerce (Recommend4You)
* Strategy Big Data : Telemarketing (Omniticket)

The variety of target sectors from various countries including Austria, Germany, Hungary, Portugal, Slovenia, Spain and Turkey will enable heterogeneous inputs for this State of the Art Analysis study as well as the OMD Project.

# State of the Art Analysis

State of the Art Analysis for each use case is studied by each company listed in alphabetical order below :

## ARD Group

### 2.1.1 Market analysis for Justice use Case

Turkey has 160.651 lawyers (as of December 31, 2021) registered to 82 Bar Associations in various cities. While some lawyers focus on working as oddsman and mediator only; more than 120.000 lawyers participate to counsel assignments in accordance with Turkish Code of Criminal Procedure (Law No. 5271) which legally requires Bar Associations, under 21 OMD FPP circumstances stipulated in the law and upon the request of the suspect or victim, to commission counsel for purposes of legal aid.

There are 7.489 and 15.614 judges among Turkey’s population of 84.680.273 people (as of December 31, 2021) according to the Turkish Statistical Institute and General Directorate of Criminal Records and Statistics.

These statistics create a great market potential and significantly increase the importance of humane value in counsel assignment systems. The number of assignments to cases for prosecution and investigation were 137.804 and 353.293 respectively for 2020.

The private sector for judicial software, particularly regarding counsels (lawyers) and Bars has a number of competitors in Turkey. One major player concentrates on search and archival of case law and jurisprudence and does not have any solutions for counsel assignment. One major player (SDD Bilgisayar Yazılım) has an established desktop software (with no mobile components) to deal with counsel assignments using a basic pointing system based on queuing and equalization of number of cases between lawyers assigned to the pool for penal procedure.

ARD GROUP, has the largest share of the market pool and has a mobile and web-based automation system for Bar Associations which employs proximity and an intelligent point-based system (depending on case type i.e. aggravated felony taking longer and having more points – differing between Bar Associations vs magistrates' court/ or “court of peace'' where cases are usually of less gruesome nature, handled more swiftly, and therefore less points are assigned). Since counsel assignment for underprivileged is a common norm throughout the modern world, OMD Project results and outputs, although currently not investigated, are bound to be applicable and suitable, albeit naturally with regulatory adjustments specific to each country, for employment and marketing to countries with a large number of lawyers/cases throughout the European Union.

### 2.1.2 Literature Survey, Definition of the Problem and Solution Approach for Justice Use Case

ARD will tackle the problem of assigning counsel to an accused person considering both attorney profiles and case descriptions. In current systems, this appointment is usually done randomly based on formal residence information and the availability of attorneys in close proximity. This often leads to the inefficient management of resources and lower satisfaction of accused persons. To perform this process in a more accurate way, an intelligent assignment model will be developed and applied under real scenarios to measure its effectiveness. ARD will address the following research questions:

* How is a better attorney assignment described? What are the parameters of an optimal assignment?
* Which data are useful to build a computational assignment model?
* Which state-of-the-art computational models are more accurate in an intelligent assignment system?
* How can such a system improve the judicial processes in real scenarios in terms of several factors including completion time, fairness and user satisfaction?

To address these questions ARD will perform R&D activities to

● describe an objective metric for ‘good’ attorney assignment,

● analyze current Optical Character Recognition (OCR) techniques to extract text from case descriptions and customize one of existing solutions for our application domain,

● design and develop a feature integration system for concurrent use of continuous, categorical, textual and even time-series data acquired from case descriptions and attorney profiles,

● design and develop an expert system for quick identification of direct assignment rules such as gender and language preferences,

● design and develop a hybrid machine learning model that will find optimal match between current case and available attorney profiles.

In general, the problem can be formally fit into “ticket routing [Xu and He, 2018]”1 or “expert recommendation [Nikzad–Khasmakhi et al. 2019]”2 problems in computer science literature, although a direct application for attorney assignment can not be found. These are usually approached as pattern classification or information retrieval tasks.

ARD will consider the problem as a similarity search, where an issue-expert pair is compared with all previously labeled/annotated pairs in an archive of completed cases. The system will predict a ‘good’ assignment score, as a label, for each pair of current issue and available attorneys. This prediction will be based on a hybrid discriminative model that will be learned from training data. Final assignment will be done after assessment of top-ranked pairs compatible with the constraints from the established expert system.

EqualityInJustice use case requires processing of personal data of attorneys as well as victims/suspects in order to enable an efficient AI based assignment solution. Therefore, data privacy and protection is a major concern for this use-case. Turkish government has released the Personal Data Protection Law (No: 6698)3 in 2016. Article 3-b clearly states the anonymization of personal data so that the personal information can not be used to identify a person directly or indirectly by combining with other information. Similarly, The General Data Protection Regulation of EU (2016/679)4 has entered into force on 24 May 2016 and applies since 25 May 2018. Recital 26 defines this issue as “data rendered anonymous in such a way that the data subject is not or no longer identifiable”.

The legal frame forces AI solution developers to a predilection of either data anonymization or incorporating synthetic data. Deloitte has drawn attention to this matter and recommended a data anonymization process in its paper ‘Preserving Privacy in Artificial Intelligence Applications through Anonymization of Sensitive Data’5 by defining the requirements, identifying personal directly identifiable information, determining the potential quasi-identifiers and finally identifying sensitive attributes in order to start the anonymization process.

Safest approach under this legal framework is to cooperate with the Union of Bar Associations and produce synthetic data, covering all information fields of the real application environment. Therefore, a major task of ARD will be developing a ‘synthetic data tool’ for this project. Expectations from EqualityInJustice by means of usability and efficiency require the most accurate information to be extracted from the synthetic data for a convenient assignment and meet the KPI’s defined in the FPP.

This approach will not entrammel the commercialization targets. Quite the contrary it will fulfill the privacy expectations of the end user with a perspective as an attorney as well as victim/suspect view.6,7,8

There will be another efficacy for this targeted synthetic data tool. It will be used a as a pluggable tool for other projects and domains of ARD.

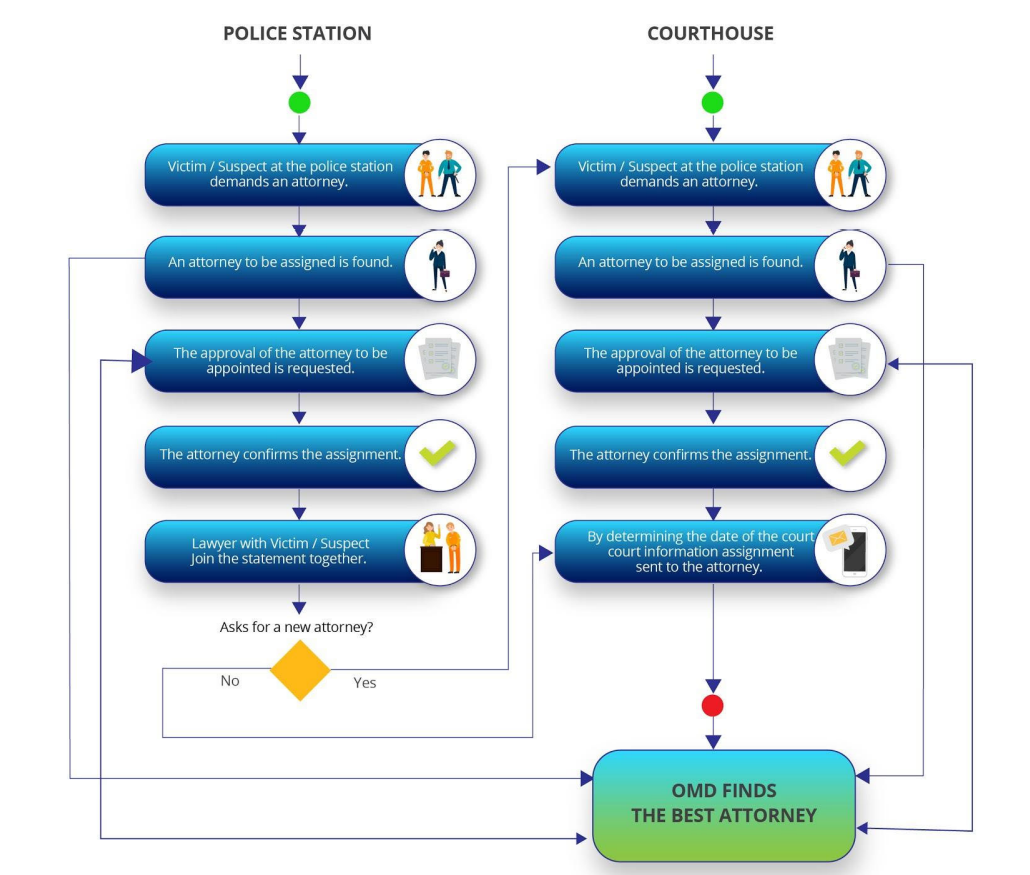


Figure 1 : Business Solution Model for Justice Use Case

## BEIA GmbH

BEIA will analyze the market for telemetry data analysis from smart industry testbeds. Beia’s use case is combining OMD with environmental IoT, for an integrated logistics that is missing. OMD platform will provide a novel and smart logistics involving adaptation to new equipments settlements and different production volumes and type of goods; investment optimization: lower time and cost with installation;plug&play sensors/AGVs system;human and robot synergy; integration of existing/new software application (ERP); applying human resources optimally based on documented individual experiences.

BEIA already supplies big data and speech processing solutions to various customers (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.

## Caretronic d.o.o.

In the OMD project, together with the consortium partners, in Caretronic we are developing a software tool that helps service providers from various fields to efficiently use resources, offer IT solutions, maintain knowledge bases, unambiguously assign each call to the right expert / provider and with the best solution to a specific problem. With the help of the OMD solution, the time and cost of operations will be reduced. By improving the overall efficiency of operations on the supply side, our project also increases customer satisfaction. For businesses that are currently using remote support at ever-increasing rates, professionals with diverse backgrounds provide support, so using an assignment system is a must. The OMD solution will increase the level of automation to exceed the state-of-the-art by developing and using a large number of artificial intelligence models, including NLP, deep learning / machine learning and a general optimization module, contributing to many sectors - healthcare, which effectively use AI models to improve the level of service .

By participating in the OMD project, our partners want to meet the strategic goals of increasing revenues and investing in research and development. In the consortium, we have complementary qualifications derived from knowledge and experience and technological capacities, and a business interest in achieving the development goals and results of the project.

## Dogus Technology

Natural Language Processing (NLP) is a field of human computer interaction. Latest development in the field using deep learning (DL) technologies raised motivation in the research. The study of dialogue systems to make machines interact with human emulating a real human conversation. There are specialized categories in the research field, such as Natural Language Understanding (NLU) and Natural Language Generation (NLG).

In our use case, our plan including related work in NLP are listed as follows:  
 · Product information and problem category will be determined based on the comments of users in natural language.  
· Responses to be given based on the problem will also be automatically retrieved from knowledge base.  
· If the problem is not resolved, integrations will be made to submit a ticket.  
· A dialog management system is planned to manage issues such as asking additional questions after the user comments when it is appropriate.  
· By using the generated training data, studies will be carried out on request/ticket classification models and asset name recognition by "NER" models [6].  
· Deep learning “DL” semantic similarity models will be developed using chat history (QA sets).

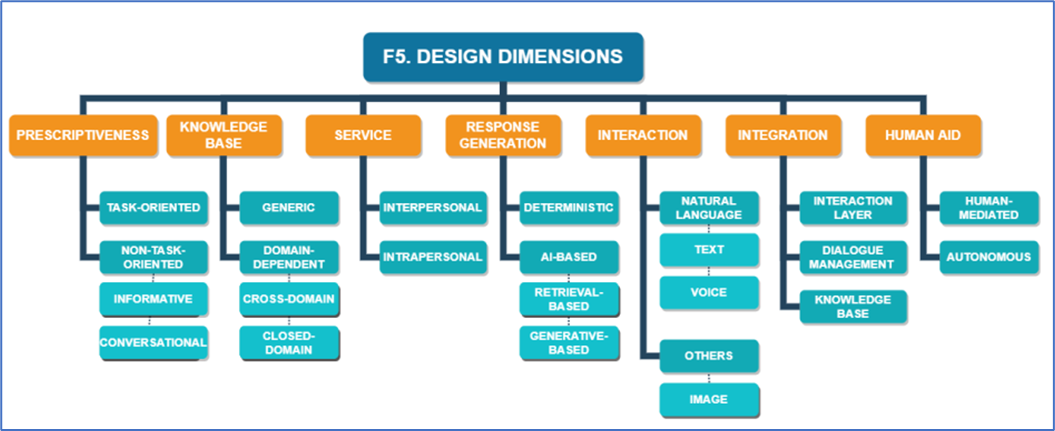


Figure 1: Design Dimensions of conversational agents [3]

Adamopoulou and Moussiades [4] and Motger et al [3] are two recent studies surveying the state of the art for dialogue systems. They also introduced a taxonomy (referred as “design dimensions”) for chatbots’ high level design specifications. Figure 1 illustrates this taxonomy.

According to [3, 4], main classification of dialogue systems is based on the design of the dialogue management system: a) “Task-Oriented”: Dialogues are limited to pre-defined tasks in a small number of domains. b) “Non-Task Oriented”: Conversations are free in which bots may answer regardless of an action. In our use case, the dialogue system refers to a Task Oriented system as it will contain an action to submit a ticket to company’s ticket management system. However, it could also be categorized as an informative non-task-oriented system due to its informative responses to user’s comments.

Our dialogue system will have a domain dependent *knowledge base*, specifically a closed domain which focuses on a single expert knowledge base rather than aiming to adapt to new contexts based on user interaction.

For *response generation*, there are two main categories, deterministic models or ML-based models (using either machine learning (ML) or deep learning (DL)). Our dialogue system will be a deterministic model retrieving adequate responses previously generated using its knowledge base and previous Q-A pair logs powered by ML and DL models.

In the project, user inputs (Figure 1 *Interaction*) will be designed over textual interaction in the interface layer. Also, it will have other parts of the dialogue management ecosystem, like a dialogue management layer and a knowledge base layer [5] (Figure 1 *Integration*).

[named entity recognition references. B-LSTM, CRF ] [6]

[text classification, intent classification, transformer, BERT referanes 7,8 ]

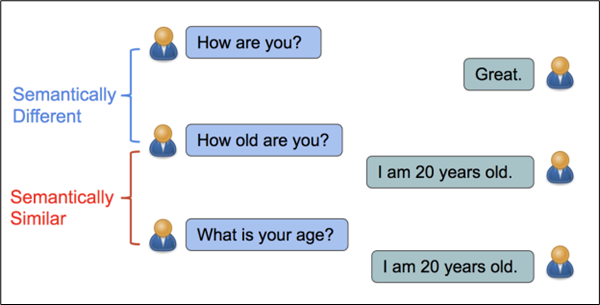


Figure 2: Semantic Similarity [9]

Since 2016, using deep learning techniques a well performing textual similarity (STS) benchmark developed and performance is still improving on recent work. In recent research in semantic similarity using dialogue data, novel unsupervised models are developed to predict question-answer (QA) pairs [9]. As highlighted in the Figure 2, similar questions could be separated to correct notions by the developed model [9]. We will use similar unsupervised STS methods to reveal QA pairs on consumer electronics real-world dialogue data, and try to observe new outcomes and achieve improvements on recent work.

Challenges

[Lack of Turkish data and noisy text]

[complexity of morphologically rich language]

Similar/Recurring Tickets Challenge: Similarity analysis techniques will be studied on existing tickets and space similarities will be removed, when a ticket is opened, the similarity analysis of this ticket will be performed with the tickets that have been opened before, and similar solved tickets will be associated with the new ticket.

Model Scaling & General Architecture Challenge: In general, OMD will be a platform where AI models invisibly optimize the operations while visibly functioning as decision support systems with advanced analytics dashboard for the management of service desks. The general OMD framework will have flexible, easy to plug & play feature AI models for different domains. Microservice architecture will be used. Therefore, especially heavy lifting computation in AI models and other computationally intensive models can easily be scaled using containers and container orchestration tools such as Kubernetes.

Complexity of customer tickets reporting one or more issues about products could be never reduced accurately to what is categorized and planned.

## Experteam

## 2.5.1. Market analysis for ITSM use case

In 2020, the top 10 IT Service Management software vendors accounted for nearly 87.2% of the global IT Service Management applications market which grew 16.5% to approach nearly $6.3 billion in license, maintenance and subscription revenues. Last year ServiceNow led the pack with a 39.6% market share riding on a 22.4% jump in ITSM licence, maintenance and subscription revenues. Atlassian was #2, followed by LogMeIn, Inc, BMC Software, and Ivanti.

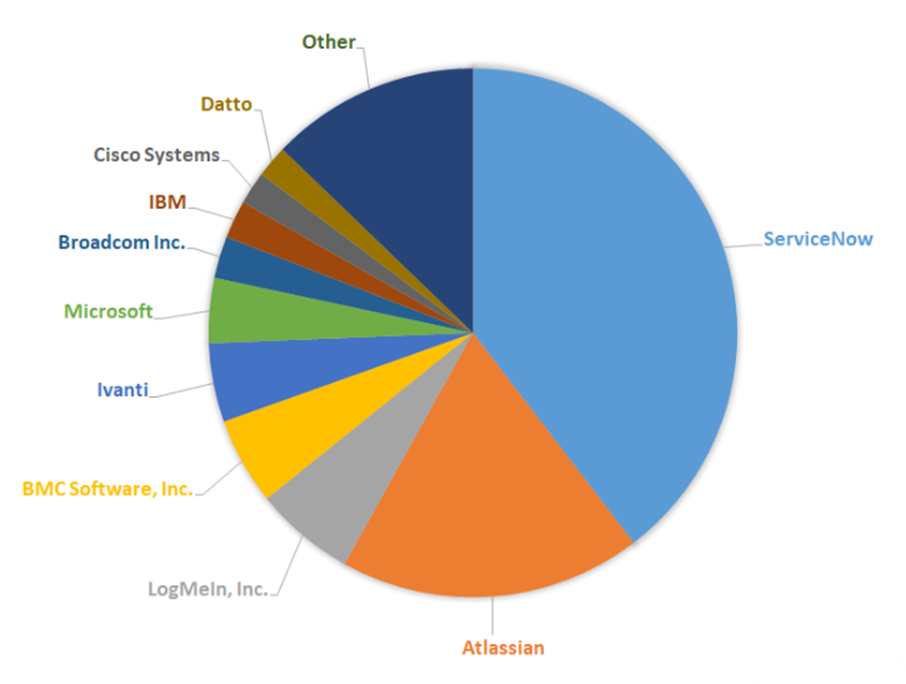


Fig. 1 – 2020 IT Service Management Applications Market Shares Split By Top 10 IT Service Management Vendors and Others

Through our forecast period, the ITSM applications market size is expected to reach $6.8 billion by 2025, compared with $6.3 billion in 2020 at a compound annual growth rate of 1.5%.

ITSM applications are being used to make a growing number of employee self-service functions possible through enterprise service management for asset, incident and project management. ITSM applications are considered a derivative market with revenue contribution to functional areas such as Project and Portfolio Management and Enterprise Resource Planning.

Massive software projects to digitise business communities fuel growth of Cloud-based vendors such as Atlassian and ServiceNow as incumbents like BMC and CA race to meet changing IT requirements.

Since our data set source is Jira, a product of Atlassian, we would like to present the companies that provide ITSM-SDM services for your information. This market can be our target market.

| **Customer** | **Industry** | **Empl.** | **Revenue** | **Country** | **Vendor** | **New Product** | **Category** |
| --- | --- | --- | --- | --- | --- | --- | --- |
| [The Home Depot, Inc.](https://www.appsruntheworld.com/customers-database/customers/view/the-home-depot-inc-united-states) | Retail | 504800 | $132.11B | United States | [Atlassian](https://www.appsruntheworld.com/customers-database/vendors/view/atlassian) | [Jira Service Desk](https://www.appsruntheworld.com/customers-database/products/view/jira-service-desk) | [IT Service Management](https://www.appsruntheworld.com/customers-database/category/it-service-management) |
| [Fiat Chrysler Automobiles](https://www.appsruntheworld.com/customers-database/customers/view/fiat-chrysler-automobiles-united-kingdom) | Automotive | 191752 | $131.64B | Italy | [Atlassian](https://www.appsruntheworld.com/customers-database/vendors/view/atlassian) | [Jira Service Desk](https://www.appsruntheworld.com/customers-database/products/view/jira-service-desk) | [IT Service Management](https://www.appsruntheworld.com/customers-database/category/it-service-management) |
| [Bosch](https://www.appsruntheworld.com/customers-database/customers/view/bosch-germany) | Manufacturing | 395000 | $87.41B | Germany | [Atlassian](https://www.appsruntheworld.com/customers-database/vendors/view/atlassian) | [Jira Service Desk](https://www.appsruntheworld.com/customers-database/products/view/jira-service-desk) | [IT Service Management](https://www.appsruntheworld.com/customers-database/category/it-service-management) |
| [Citibank National Association](https://www.appsruntheworld.com/customers-database/customers/view/citibank-national-association-united-states) | Banking and Financial Services | 239000 | $76.40B | United States | [Atlassian](https://www.appsruntheworld.com/customers-database/vendors/view/atlassian) | [Jira Service Desk](https://www.appsruntheworld.com/customers-database/products/view/jira-service-desk) | [IT Service Management](https://www.appsruntheworld.com/customers-database/category/it-service-management) |
| [T-Mobile Retail USA](https://www.appsruntheworld.com/customers-database/customers/view/t-mobile-retail-usa-united-states) | Communications | 75000 | $68.40B | United States | [Atlassian](https://www.appsruntheworld.com/customers-database/vendors/view/atlassian) | [Jira Service Desk](https://www.appsruntheworld.com/customers-database/products/view/jira-service-desk) | [IT Service Management](https://www.appsruntheworld.com/customers-database/category/it-service-management) |

We present to your information the developments experienced in the products of the top 10 companies that stand out according to their market shares in the sector as of 2020-2021, with the below table.

2.5.2. Sota for ITSM use case

| **Rank** | **Vendor** | **YoY Growth** | **Recent Developments** |
| --- | --- | --- | --- |
| 1 | [ServiceNow](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/servicenow?apps=true) | 22.4% | ServiceNow expects to have an estimated $5.5 Billion in revenues worldwide for CY2021, and an international presence that can generate new business across geographies. ServiceNow acquired the following companies in 2021 to accelerate innovation and enhance talent: DotWalk (Nov 2021), Gekkobrain (Oct 2021), Mapwize (Aug 2021), Swarm64 (Aug 2021), Lightstep (May 2021), Intellibot (Mar 2021), ElementAI (Jan 2021). |
| 2 | [Atlassian](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/atlassian?apps=true) | 11.6% | The company stopped selling new on-prem server licenses as of 2021. While the company is supporting existing on-prem server customers until 2024, the idea is to now move them to the cloud and this offering should help. One thing that is clear is that the pandemic has accelerated the move to the cloud by companies of every size, and this should encourage the company’s largest customers to make the move.Atlassian’s number of total customers increased to 236118 as of June 2021 from 174,097 at June 30, 2020. |
| 3 | [LogMeIn, Inc.](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/logmein-inc?apps=true) | 30.0% | In 2020, LogMeIn has launched Remote Deployment for GoToMyPC enabling IT administrators and business professionals to remotely deploy, install, and configure GoToMyPC remote access software across any number of computers simultaneously. |
| 4 | [BMC Software, Inc.](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/bmc-software-inc?apps=true) | 12.7% | BMC Software has recently announced the industry’s first end-to-end integrated ITSM and ITOM platform which brings together the capabilities of ITSM and ITOM into BMC Helix. Integrated solutions will enable IT teams to discover, optimize, remediate and deliver through an omnichannel for IT and business users. Such solutions will eliminate silos to make better-informed decisions and enhance the customer experience. It’ll also help in discovering unknown assets in multi-cloud, monitor & predict events, uncover & remediate security vulnerabilities and provide 360-degree intelligence environment. |
| 5 | [Ivanti](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/ivanti?apps=true) | 0.2% | Ivanti’s software is utilized by over 40,000 customers across various industries and five continents. From patch management and IT security solutions to IT Asset Management, IT Service Management, and IT Systems Management to solutions for the warehouse, Ivanti changes the way businesses work. |
|  |  |  | In 2021, Ivanti acquired Cherwell and their Cherwell Service Management platform to provide even greater capabilities in ESM and also in 2021, released Ivanti Neurons for HR and Ivanti Neurons for Facilities to provide the same consistent experience for employees across their onboarding, remote work services, and return to the office initiatives. |
| 6 | [Microsoft](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/microsoft?apps=true) | 14.3% | GitHub, which has 73 million developers, has emerged as the focal point of Microsoft’s ITSM strategy. Customers are choosing GitHub Enterprise to provide their developer teams the right platform to build, ship, and maintain software. In 1QFY22, it introduced more than 70 enterprise features like GitHub Actions for developers to better manage their workflows. |
| 7 | [Broadcom Inc.](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/broadcom-inc?apps=true) | 5.4% | Broadcom Inc. has acquired CA Technologies to build one of the world’s leading infrastructure technology companies. The all-cash transaction represents an equity value of approximately $18.9 billion and an enterprise value of approximately $18.4 billion. CA Technologies has more than 90000 customers across the globe. |
| 8 | [IBM](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/ibm?apps=true) | -8.8% | ITSM remains a key focus for IBM after spinning off units like AppScan, Big Fix, Notes and Domino to HCL. Its Red Hat purchase could be a boon for IT security. |
| 9 | [Cisco Systems](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/cisco-systems?apps=true) | 6.7% | Cisco ACI, Cisco Cloud Center, and ServiceNow together automate and orchestrate service provisioning & activation workflow by completely hiding complexities of infrastructure from IT or LOB end-users through software-defined abstraction. Using this powerful integrated solution, IT organizations can achieve accurate service mapping and extremely fast service provisioning, while having automated coordination between dynamic Hybrid IT infrastructure and relevant business services. |
| 10 | [Datto](https://www.appsruntheworld.com/cloud-top-500-applications-vendors/datto?apps=true) | 9.7% | Datto protects business data and provides secure connectivity for tens of thousands of the world’s fastest-growing companies. Datto’s Total Data Protection solutions deliver uninterrupted access to business data on-site, in transit, and in the cloud. Thousands of IT service providers globally rely on Datto’s combination of pioneering technology and dedicated services to ensure businesses are always on. |

When we examine the above-mentioned companies and their software, the most important innovation we reveal within the scope of OMD will be the scoring technique we will use in the problem of assigning the ticket to the most accurate expert, and accordingly our optimization technique. With this aspect, we perform assignments with a more precise accuracy than our competitors in the market.

## FrontEndART Software

### Market analysis for Optimal software maintenance task assignment use case

Nowadays, there are only a very few fields where no software is used. Practically, almost any part of our modern life is related to or relies on software products. Thus, it is crucial that bugs in these software are corrected as soon as possible. Sometimes it is not only a requirement for keeping the reputation of the vendor, but also a must due to SLAs and other contracts.

But who should correct a bug? In a large company that handles hundreds of software modules it is important to optimize the overall task assignment process in order to perform efficiently and effectively. Although it might seem to be simple (e.g. assign the task to the developer who can fix it most quickly), it is not. Experience, technical knowledge, familiarity with the code, number of other tasks already assigned to the developer and several other factors can influence the decision. A good assignment strategy can speed up the overall bug fixing performance of the developer team.

There are more than 9,000 software companies in Europe, and more than 4,000,000 professional developers just in the top 10 European countries. The software market revenue was more than €135 billion in 2021, and is expected to grow over €190 billion in 5 years. Even if optimal task assignment can save only a small portion of this, that means a large saving.

### State of the art for Optimal software maintenance task assignment use case

Optimal task assignment for finishing the last task as soon as possible is a well known mathematical problem. However, the theoretical/mathematical problem does not really cover software maintenance task assignment as there are several factors that influence the outcome [Capretz and Ahmed 2010], and there can be several objectives that have to be fulfilled.

Automatic task/bug assignment is an active research field [Gadge and Mangrulkar 2017], [Panichella et al. 2013]. There are several methods that concentrate on different aspects and objectives of this problem. Some methods may concentrate on the problems related to assigning bugs reported by humans in natural language [Shokripour et al. 2015] where the location of the bug in the source is mostly unknown at assignment time. Others consider only the developers’ bug fixing and development activities [Khatun and Sakib 2018] or try to mitigate the effect of developer rotation of the organization [Etemadi et al. 2022], and there are algorithmic solutions that take into account several factors [Tang 2020]. Furthermore, most complex automatic task assignment methods do not deal with the overall picture; they usually assign the bugs individually, which can result in a very unbalanced assignment (where one developer has most of the bugs while other developers have no tasks).

In our case the maintenance tasks are reported by static analysis tools, which means that the nature of the problem is well classified and the location of the code to be enhanced is also known. Although there are several methods using AI techniques for bug assignment [Jonsson et al. 2016] as far as we know, no solutions employed AI techniques in the context of automating optimal task assignment among a given set of developers.

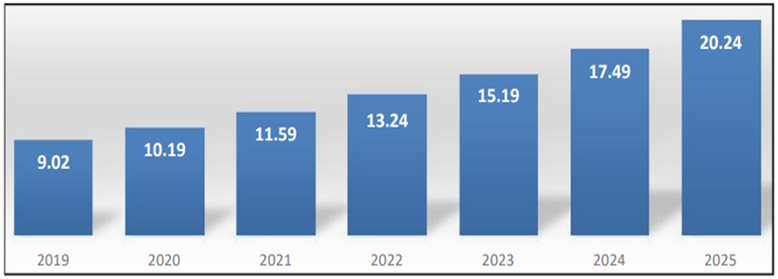
## FTP-LDA &ISEP

2.7.1. Market analysis for e-commerce use Case

Over the past years, there has been a profound change in population buying habits. The pandemic era led to a greater demand and adoption of B2B e-commerce. For other side as the population increases so does the marketplace, leading to a demand for a better and differentiated offer. E-Commerce has completely changed the way we shop; customer expectations have increased, and an appropriate response needs to be built.

The ability to check out the product in its almost real-life appearance is an increasingly necessity to capture our target customers and promote online purchases. Innovation through optimization on e-commerce platforms, improves customer relation.

The e-commerce market evolves the sale of all type of products and services through online channels. The main model types are “Business to Business” (transactions between businesses, like manufacturer to wholesaler) and “Business to Consumer”, where the global market reached a value of approximately 13 billion euros (European definition of billion) in 2021. The recent years percentage growth and the forecast of the next years can be seen in the figure below.



International Journal of Applied Marketing and Management (Volume 6 Issue 1 - 2021)

Markets expect to invest more in e-commerce innovation and technology to boost the overall market development. The approach of a channel like this can provide several opportunities for business segments and companies are investing to expand their growth and reach consumer needs. The geographical market share is divided in 40.92% for Asia & Pacific, 24.99% for North America, 22. 45% for Europe, 6.19% for Middle East & Africa and 5.45% for Latin America.

The increased availability of internet, electronic devices and unexpected realities that recently affected the world globally (like COVID-19 pandemic) had a meaningful impact across the e-commerce market and retail value chains, where consumer expectations and behaviors had seen a recurrent evolution. In today’s scenario, the users give a lot of importance to ease of use, speed and intuitive experiences across different devices and channels.

E-commerce businesses are efficient, reachable and profitable, although highly fragmented and with strong competition. Therefore, companies are available to incorporate several strategies in order to sustain their presence.

### 2.7.2. State of the art for e-commerce use case

As more people have access to the internet, the amount of available information also increases. The companies have taken advantage of this reality and started to export their business to the web, being called E-commerce. The increased information on the internet has lots of advantages and disadvantages. The main disadvantage is when someone is making a simple search on the internet, the amount of results makes it difficult to choose.

Be aware of this problem, under the scope of OMD (Optimal Management of Demand) project the Portuguese consortium has the objective to develop a Chatbot to E-commerce scenario.

The development of Conversation/Chat Agents or chatbots is not a new thing, in fact it has increased over the last years [3]. A chatbot is a system that allows a conversation between a human and the machine [4]. To allow such conversion lots of techniques and methods have been created. The methods and techniques that have been created are in the artificial intelligence area and the algorithms used to train it are machine learning and deep learning algorithms [5]. ML (Machine Learning) is one of the Artificial Intelligence areas. This consists in the study of algorithms that allows the machine to learn and improve from experience without the need to program it [6]. NLP (Natural Language Processing) is the process that studies and gives the machine the ability to understand human speech [7]. Some of the techniques used in the NLP area are NER (Name Entity Recognition), SA (Sentiment Analysis) and Intent Classification.

### 

### 

### 

### 2.7.2.1**.** Tools to create chatbots

Currently, it is possible to use tools that train/create models and even prepare data to achieve a chatbot. Some of these are:

● BERT - Due to the shortage of trained data and the fact that NLP is a big area, the current task-specific datasets contain few data. The problem is that modern deep learning-based NLP models benefit from large datasets and training huge datasets takes a lot of time, time that is sometimes expensive. To solve this problem, researchers have developed a model called BERT (Bidirectional Encoder Representations). This new model can be fine-tuned using small datasets and, besides this, showed an improved accuracy compared to training from scratch a model [11].

● Spacy - is a free, open-source library for NPL in the language Python. Like BERT, Spacy allows to train models, text classification, serialization, and many more to understand large volums of text [12]. The features it provides are Tokenization, POS(Part-of-speech) tagging, Dependency parsing, Lemmatization, SBD (Sentence Boundary Detection), NER, EL (Entity Linking), Similarity, Text Classification, Rule-based Matching, Training and Serialization.

● Rasa - is an open-source machine learning framework. This framework has the objective to allow a machine to understand messages, allowing the machine to have conversations and the ability to connect to messaging channels and even to APIs [13]. On this framework, it is possible to create stories that are a must in the development of chatbots.

● UBIAI - is a text annotation tool for NLP. In this tool it is possible to upload documents where it is possible to classify it for then use it for training models. It is also possible to train models [14].

● Hugging Face - is an open-source platform that provides tools that allow the training and deployment of machine learning models. Their tools can be used from an API [15]. Hugging Face has it's own NLP library called Transformers witch supports BERT model, ALBERT and many more.

### 2.7.2.2**.** Related Works

In literature it is possible to find examples of several works that describe the development of chatbots, below we will give examples of some. We will conclude this section comparing the different works in terms of incorporation of NLP models, acting like recommendation system, integration with other systems and incorporation or not of a database

A chatbot, called Hebron, was developed to help the members of the Covenant University Community to save time when going to CUSM (Covenant University Shopping Mall). The time problem is the fact that CUSM doesn't have an online inventory, so when a member goes there to buy some products and the store doesn't have it, the member waists time to go there. The developed chatbot could be asked if the desired product is available and perform the product's payment, making all the process less time-consuming. The technologies used in the development of the chatbot were Reacts.js to build the interface, MySQL to store and send to the user the list of items and the quantity and Spacy was used for the ML (Machine Learning) part [8].

The negotiation of the price of a product is an important factor to a business, it has the consequence of increasing the client’s satisfaction. A chatbot was created to allow negotiation of price in an e-commerce scenario. The proposed methodology presented to develop this chatbot was getting the query from the user, dialogue roll-out, sentiment analysis and natural language processing [9].

When a person wants to buy something on the internet, they will waste lots of time browsing and searching in which e-commerce is the best to buy that same product. To solve this time-consuming search, a chatbot was developed. The framework used to develop this chatbot was RASA due to the number of modules available [4].

Accessing a college website, if a person doesn't study/work there, when searching for desired information it will be difficult for that person. A chatbot is proposed to help a person obtain the desired information, this information can be about admission, examination and many more. To implement such a chatbot, AIML files were used to store the question-and-answer pairs, WordNet to extract information from the given text, Semantic Sentence Similarity to compare a question given to the already existing ones. Finally, a log file was used to store the questions that the chatbot couldn't answer [10].

Due to the lack of chatbots to help students in the search of hostels with recommendations systems that satisfy their needs. Having knowledge of this problem, a chatbot was made to solve it [16]. This chatbot comes in the form of an API, making it easy to be integrated in different systems. This API was made using PHP and Flask. The chatbot was made using NLTK and Spacy to process the text received. A database was also made to allocate the list of hotels using MySQL.

Starting the studies in the university, their life will change due to the different environment it is compared to the previous environment. This new environment will cause lots of stress to new students due to the lack of guidance about bureaucracy and information about the university administrative process and general questions. To ease this stress, a chatbot, called S.A.N.D.R.A., was made for the Brazilian public university [17]. In the development of the chatbot, a dataset of all the FAQ in the university website and a database containing all the 1.453 subjects. To avoid errors in the users, query the distance of Levenshtein was used. To process the user’s query, Scikit-learn and BERT were used.

As more and more users decide to do their shopping on the web, the need to market business products online increases, but not every company is doing it. To help the company’s market their products online and save them money by replacing some human tasks with informatization, a chatbot was made [18]. This chatbot was made using ManyChat integrated in a Facebook fan page.

Due to the recent waves of COVID-19 and the fact that exists data for the Vietnamize language in NLP area, a chatbot was made for QA (Questions and Answers) for COVID-19 information in Vietnamize. To develop this chatbot BERT and ALBERT model were used for the implementation of a QA tasks, for the framework used RASA was chosen [19].

### Table 2 - Chatbots Characteristics

| Chatbots | Has an API that can be integrated in another system? | NLP models | Recommendation system? | Connected to a database? |
| --- | --- | --- | --- | --- |
| [8] | No | Spacy | No | Yes |
| [9] | No | Nan | No | No |
| [4] | No | Rasa | No | Yes |
| [10] | No | WordNet | No | Yes |
| [16] | Yes | NLTK and Spacy | Yes | Yes |
| [17] | No | Scikit-learn and BERT | No | Yes |
| [18] | No | ManyChat | No | Yes |
| [19] | No | BERT, ALBERT and Rasa | No | No |

## Strategy Big Data

NLP-based technologies are supporting contact centers and telemarketing services in many different ways. Applications generally focus on processing calls, text and also success KPIs (number of sales, recurrence...) by identifying specific characteristics related to an agent, a customer and a product / campaign. The main challenge is to leverage this background to improve the success rate of a new customer not previously contacted and without a basic profile generated. Some socio-economic characteristics of a potential new customer could be available in advance (i.e. income levels, cultural level, social...) based, for example, on location / IP address. This could lead to an association to predict the level of rapport that increases the likelihood of customer acquisition.

To materialize these innovations, SBD brings to the project different technologies for the generation of customer profiles, agents and rapport prediction, highlighting:

SBD will develop new models for purchase prediction from the information generated in the components:

- AI Geo-Demographics Profiler

- Mood/Cultural/Socioeconomic enhanced customer profiler

- AI Agent Enhanced Profiler

These models will be integrated into the component (Rapport Prediction-based Call generator). To do so, SBD will carry out an exhaustive analysis of the state of the art in profiling and prediction of buying potential. Predicting a set of customers who are more likely to accept an offer or a sale based on their personal characteristics or buying behavior is a hot research topic.

Palaniappan et al. proposed a classification approach for customer profiling that highlights the better results of the decision tree compared to Naïve Bayes and Random Forest. These techniques can be used when there is sufficient customer information. In addition, the objective is to generate a basic profile for a new customer without prior registration.

SBD will develop new models for socio-cultural profiling of potential customers by means of a priori profiling of customers based on external sources and the application of new AI-based techniques for the analysis of paralinguistic voice descriptors, applied to the conversation of customers and agents.

These models will be integrated into the components:

- AI Geo-Demographics Profiler

- Mood/Cultural/Socioeconomic enhanced customer profiler.

For its development, SBD will carry out a state-of-the-art analysis of current techniques and proposals. Regarding socioeconomic level prediction based on demographic and geo-referenced data, Ren et al. provided a model trained with online and offline requirements extracted from geo-referenced datasets to predict socioeconomic levels of urban regions. Kianmehr et al. proposed to predict socioeconomic characteristics based on IP address geolocations. Abitbol et al. proposed a methodology to infer socioeconomic level based on numerous sources: Twitter, census data, LinkedIn, and Google Maps.

SBD will develop a new and innovative Rapport (affinity) prediction model, realizing the best association between agent, customer and product. This model will be integrated in the "Rapport Prediction-based Call generator" component, and will offer new services (based on existing network analysis and customer behavior) to add the trend and probability of product purchase.

To this end, SBD will conduct an analysis of the current state of the art. The link between mood and buyer tendency has been extensively studied.[1][2][3]. In addition, the relationship between the customer and the seller, as well as the socioeconomic affinity between the two are key factors.[4][5]. Systematic and large-scale analysis of mood could suggest collective emotional trends for existing social and economic indicators. [6]. The findings [7] show that large social events have a relevant effect on various dimensions of public mood (i.e., tension, depression, anger, vigor, fatigue, confusion). Bann et al.[8], proposed a commensurate approach to how people convey their understanding of emotions through the language they use.

It has been shown that shared social environment, similarity in educational levels, and social influence together lead socioeconomic groups to exhibit stereotyped behavioral patterns, such as shared political opinions [9] or similar linguistic patterns[10]. In addition, mechanisms associated with parasocial interaction and relationships will be taken into account. Hartmann et al.[11] suggests that interactive encounters can still be called parasocial. Therefore, measurement tools and scales will be analyzed. [12]. In addition to the analysis of customer-agent linkages, the analysis of buyer trends will be integrated as a parameter to be taken into account. [13][14] will be integrated as a parameter to be taken into account.

## University of Szeged

### Literature review and state-of-the-art in NLP

#### General introduction to NLP

Language can be defined as a structured system of communication, where the structure of the language is its grammar and the free components are its vocabulary. Natural Language Processing (NLP) is devoted to making computers understand the statements and words written in human languages, with a combination of computational linguistics, statistical, machine learning, and deep learning models. NLP is a subfield of applied Artificial Intelligence (AI) that focuses on understanding, interpretation, analysis, manipulation and generation of natural language data by machine learning and most often deep learning methods. NLP can be divided into two essential parts: Natural Language Understanding (NLU) and Natural Language Generation (NLG). NLU analyzes natural language, while NLG can produce meaningful sentences and paragraphs.

Practical applications of NLP cover the whole spectrum of text-related tasks, including, among others, different analysis methods such as morphological analysis, which is the process of determining the morphemes from which a given word is constructed. Semantic analysis methods take into account the greater context and try to capture the meaning of the given text. Various classification tasks are available including sentiment analysis which is used to determine whether the emotional tone of a text is positive, negative or neutral, or other document categorizations to separate text based on their subjects. Machine translation refers to the process of automatically translating content from one language to another. Automatic question answering systems can automatically answer a question, in the same way as humans do, while information retrieval systems deal with the organization, retrieval, and evaluation of textual information from document repositories.

Basic NLP tools are also available for language analysis and processing. Dependency parsing is a procedure to examine the dependencies among the constituent phrases and parts of a sentence. Sentence segmentation is the process of identifying sentence boundaries between words in different sentences. Tokenization approaches separate a piece of text into smaller units called tokens based on either rules or advanced AI methods. Two further related approaches are stemming that reduces a word to its word stem, and lemmatization that converts a word to its root. A widely developed and applied field of NLP tools is Named Entity Recognition, which is the process of detecting and classifying arbitrary types of named entities in the text. This above representative list is far from being complete, there are many more other approaches that can now be used more-or-less routinely by utilizing API-based libraries such as SpaCy1 in the Python language.

#### State-of-the-art in text embeddings and representation

The most basic step for the majority of NLP tasks is to convert words into numbers. One essential and also a broad field of NLP applications is to produce meaningful embeddings which can be used to convert arbitrary text input to a machine comprehensible form. Word embeddings are one of the most popular representations of documents, these are a type of word representations that allows words with similar meanings to have a similar representation. There are many different techniques to make such representation, which have become more and more sophisticated with the rapid development of the field.

Traditional techniques rely on simple bag-of-words like representations, where a vocabulary of the unique tokens is defined for the target corpus, i.e., a collection of documents. The easiest way to create representations is to use a lookup table. In this case, first, a dictionary is created from a large amount of text, to reduce the dictionary, only the roots of the words are included. Then a unique number will be assigned to each word, which will be the representation of the word. The disadvantage of this method is that all the out-of-vocabulary words will get the same identifier, and words with similar meanings will get different representations. Another approach is the One-Hot Encoding method, which assigns a dictionary-sized sparse binary vector to each word. For that word, all values ​​in the associated vector are zero, except for the dictionary sequence number associated with the word. In these methods, the word representation does not contain any semantic information of words.

The count-based as well as the somewhat more advanced statistics-based representations give a slightly better representation. The most commonly used approach is the term frequency-inverse document frequency (TF-IDF) weighting approach, which takes into account the relevance of a given word to a document in the collection of documents, i.e., the corpus. These methods produce large, usually sparse and binary embedding vectors. More advanced representation techniques can reduce the dimensionality of the embeddings by applying a shallow neural network in an unsupervised setting, and produce dense real-valued vectors which also encompass context information to some extent, i.e., the embeddings are semantically meaningful. These representation techniques rely on the distributional hypothesis, which states that words occurring in the same contexts tend to have similar meanings. That is, we can characterize the meaning of a word based on its context in a given corpus. One of the most cited and widely used embedding techniques is the word2vec method2,3, which exists in many different variations depending on the domain of use and the properties of the underlying task and input at hand. Word2vec models can be trained using two algorithms: Skip Gram (SG) and Common Bag Of Words (CBOW). The CBOW model takes the surrounding context of a word as input and tries to predict the word based on the context. The input of SG is the target word itself, and the algorithm tries to use this word to predict its neighbors. When generating the representation, word2vec takes into account only close context, whereas the GloVe (Global Vector)4 approach takes into account the entire corpus. Training GloVe is performed on aggregated global word-word co-occurrence statistics from a corpus, which is an unsupervised learning algorithm to obtain vector representations for words. Both word2vec and GloVe approaches can establish the semantic similarity between words, but their disadvantage is that they cannot deal with the words of the same form which are derived from different contexts.

The next big step towards solving this issue was ELMo (Embeddings From Language Models)5, which, depending on the context, dynamically produces the corresponding word embeddings. The model consists of stacked CNN (Convolutional Neural Network), BiLSTM (Bidirectional Long-Short Term Memory) and Embedding layers. As a result, the representations of the word in different contexts can be different.

Currently, the state-of-the-art in modern NLP use cases is to use large Transformer-based deep neural networks6 which undoubtedly made a large leap in NLP research and application. Introduced by Google in 2017, Transformers consist of stacked layers of encoder-decoder architectures together with attention and self-attention layers. These neural networks are among the newest and currently the most powerful classes of models with applications not only in the field of NLP but also in image processing7. The Transformer neural network aims to solve sequence-to-sequence tasks while handling long-range dependencies with ease due to the attention architecture. The essential benefit of these models is their strong ability to capture the semantics of words in context via rich contextualized embeddings. The models employ either Encoder and/or Decoder layers to encode the textual input and provide a large dimension dense embedding vector. Multiple variants and families of Transformers have been published and although they are developed for a more or less well-defined set of specific NLP tasks, the common aim for all is the general usability with as less further tailoring of model weights and architecture as possible. The models are pre-trained on large datasets in an unsupervised fashion using language modeling tasks with the aim of providing the best possible linguistic representation. Using these models essentially reduces the need for expert-defined feature spaces and tailoring specific model architectures, and solves the issue of lacking large annotated datasets at the same time.

The most cited Encoder-based language model is BERT8 (Bidirectional Encoder Representations from Transformers), while the GPT9 (Generative Pre-trained Transformer) family of models is the Decoder-based counterparts. These models were pre-trained on a large corpus in a self-supervised fashion, which means that there was an automatic generation of inputs and labels from those texts.

BERT is a deeply bidirectional, unsupervised language representation technique, pre-trained using only a plain text corpus. Pre-training was performed using the Masked Language Model (MLM) and Next Sentence Prediction (NSP) tasks. MLM tries to predict a set of randomly masked words in the corpus, while NSP tries to predict if sentence A comes after sentence B, which inherently teaches BERT to understand longer-term dependencies across sentences. In general, as a consequence of its structure and training strategy, BERT is expected to be better for language understanding tasks, such as classification or extractive information retrieval. Following its first publication, many new alternatives have been developed inspired by the original BERT architecture10. They are pre-trained for different languages, i.e., huBERT for Hungarian, Chinese-BERT for Chinese or BETO for Spanish, or optimized for a particular domain such as SciBERT for scientific text, BioBERT for the biomedical domain or ClinicalBERT for clinical text.

As mentioned above the Decoder-based Transformer model is GPT, which excels in text generation and related tasks. There are many model variants for the GPT family. The first version of GPT contained 117 million parameters, the second version, GPT-2 has around 1.5 billion parameters, while the largest member of the latest version, the full GPT-3 has over 175 billion parameters. The basic idea behind the ever increasing architecture of the Transformer models is to exploit the great flexibility and variability provided by the huge dimensionality of the parameter space to encode as much information from a large general corpus as possible during the unsupervised pre-training phase11. Subsequently, when a specific downstream task with a more focused domain is considered, the general model can be fine-tuned to this particular task using a few thousand or even hundreds of labeled examples. Cutting edge models, such as variants of GPT-3 with their billions of parameters can even drop the fine-tuning part and apply the model directly (zero-shot), by showing a single (one-shot), or a few (few-shot) labeled examples without updating the weights of the language model12. In the different X-shot settings the user provides descriptive samples to the model where the input and the required output formats are clearly defined. This is called a prompt, and the field of producing good quality prompts is called prompt engineering. The performance of the models heavily depends on the quality of the prompts, and basically the ability of the model to understand the task at hand relying on the provided samples.

Briefly summarizing the above it can be concluded that in recent years, pre-trained language models have become the de facto state-of-the-art workhorse in most if not all downstream NLP developments and applications. These models are pre-trained with the general aim in mind to generate better contextualized representations of text. Pre-trained models have the potential to address most of the limitations of the earlier representation methods, however, this great performance comes with a price because training a multi-billion parameter model requires a huge corpus and significant computing power. However, with optimized and creative fine-tuning approaches, the models can easily be tailored to solve most downstream tasks in NLP.

# References

# ARD Group

*1.* *Xu and He, (2018), Expert recommendation for trouble ticket routing, Data & Knowledge Engineering 116, 205–218.*

*2.* *Nikzad–Khasmakhi et al., (2019), The state-of-the-art in expert recommendation systems, Engineering Applications of Artificial Intelligence 82, 126–147.*

*3.* *Turkish Republic Personal Data Protection Law, No : 6698, Official Newspaper Date 07.04.2016 No: 29677, Article 3, 4 and 5.*

*4.* *The General Data Protection Regulation (GDPR) 2016/679 of EU, Recital 26.*

*5.* *Deloitte, Preserving Privacy in Artificial Intelligence Applications through Anonymization of Sensitive Data, Issue 02/2022, pp 5-6*

*6.* *The Alan Turing Institute, James Jordon, Florimond Houssiau, Giovanni Cherubin, Samuel N. Cohen, Lukasz Szpruch, Mirko Bottarelli, Carsten Maple, Adrian Weller, Synthetic Data – what, why and how?, May 2022, pp 7,8,9*

*7.* *Harrison Wilde, Jack Jewson, Sebastian Vollmer, Chris Holmes, Foundations of Bayesian Learning from Synthetic Data, 2021, pp 2,3,4*

*8.* *Mikel Hernandez, Gorka Epelde, Andoni Beristain, Roberto Alvarez, Cristina Molina, Xabat Larrea, Ane Alberdi, Michalis Timoleon, anagiotis Bamidis, Evdokimos Konstantinidis, Incorporation of Synthetic Data Generation Techniques within a Controlled Data Processing Workflow in the Health and Wellbeing Domain, MDPI, Electronics 2002, 11, 812, pp 4-7*

# FEA

1. *Gadge, T. S., & Mangrulkar, N. (2017). Approaches for automated bug triaging: A review. In proceedings of the 2017 International Conference on Innovative Mechanisms for Industry Applications (ICIMIA), Bangalore, 2017, (pp. 158-161), doi: 10.1109/ICIMIA.2017.7975592*
2. *Shokripour, R., Anvik, J., Kasirun, Z. M., & Zamani, S. (2015) A time-based approach to automatic bug report assignment. Journal of Systems and Software, Volume 102, 2015, (pp. 109-122), doi: 10.1016/j.jss.2014.12.049*
3. *Khatun, A. & Sakib, K. (2018). A Bug Assignment Approach Combining Expertise and Recency of Both Bug Fixing and Source Commits. In Proceedings of the 13th International Conference on Evaluation of Novel Approaches to Software Engineering - Volume 1: ENASE, 2018, (pp. 351-358), doi: 10.5220/0006785303510358*
4. *Jonsson, L., Borg, M., Broman, D. et al. Automated bug assignment: Ensemble-based machine learning in large scale industrial contexts. Empirical Software Engineering 21, 2016, (pp. 1533–1578), doi: 10.1007/s10664-015-9401-9*
5. *Etemadi, V., Bushehrian, O., Robles, G. (2022). Task assignment to counter the effect of developer turnover in software maintenance: A knowledge diffusion model. Information and Software Technology, Volume 143, 2022, doi: 10.1016/j.infsof.2021.106786.*
6. *Tang. F. (2020) Optimal Complex Task Assignment in Service Crowdsourcing. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Pages 1563-1569. doi: 10.24963/ijcai.2020/217*
7. *A. Panichella, B. Dit, R. Oliveto, M. Di Penta, D. Poshynanyk and A. De Lucia. (2013) How to effectively use topic models for software engineering tasks? An approach based on Genetic Algorithms. 2013 35th International Conference on Software Engineering (ICSE), 2013, pp. 522-531, doi: 10.1109/ICSE.2013.6606598.*
8. *L. F. Capretz and F. Ahmed. (2010) Making Sense of Software Development and Personality Types. in IT Professional, vol. 12, no. 1, pp. 6-13, Jan.-Feb. 2010, doi: 10.1109/MITP.2010.33.*

# USZ

1. <https://spacy.io>
2. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. arXiv:1301.3781
3. Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111–3119).
4. J. Pennington, R. Socher, C. D. Manning (2014). GloVe: Global Vectors for Word Representation. Empirical Methods in Natural Language Processing (EMNLP), 1532-1543.
5. M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer (2018). Deep contextualized word representations. arXiv:1802.05365
6. Vaswani, A, Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., and Polosukhin, I. (2017). Attention Is All You Need, arXiv:1706.03762.
7. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, N. Houlsby (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv:2010.11929
8. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K. (2018). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, arXiv:1810.04805v2
9. Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D.M., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., Amodei, D. (2020). Language Models are Few-Shot Learners, arXiv:2005.14165
10. X. Qiu, T. Sun, Y. Xu, Y. Shao, N. Dai, X. Huang (2021). Pre-trained Models for Natural Language Processing: A Survey. arXiv: 2003.08271
11. Han, X., Zhang, Z., Ding, N., et al. (2021). Pre-trained models: Past, present and future. *AI Open*, **2**, 225. DOI: 10.1016/j.aiopen.2021.08.002
12. Kalyan, K.S., Rajasekharan, A., and Sangeetha, S. (2021). AMMUS : A Survey of Transformer-based Pretrained Models in Natural Language Processing, arXiv: 2108.05542

# SBD

1. Bambauer-Sachse, S., & Gierl, H. (2009). Can a positive mood counterbalance weak arguments in personal sales conversations?. Journal of Retailing and Consumer Services, 16(3), 190-196.
2. Liljander, V., & Mattsson, J. (2002). Impact of customer preconsumption mood on the evaluation of employee behavior in service encounters. Psychology & Marketing, 19(10), 837-860
3. Hume, M., & Mort, G. S. (2010). The consequence of appraisal emotion, service quality, perceived value and customer satisfaction on repurchase intent in the performing arts. Journal of Services Marketing.
4. Skinner, H., & Stephens, P. (2003). Speaking the same language: The relevance of neuro‐linguistic programming to effective marketing communications. Journal of Marketing Communications, 9(3), 177-192.
5. Nancarrow, C., & Penn, S. (1998). Rapport in telemarketing‐mirror, mirror on the call?. Marketing Intelligence & Planning.
6. Pilato, G., & D’Avanzo, E. (2018). Data-driven social mood analysis through the conceptualization of emotional fingerprints. Procedia computer science, 123, 360-365.
7. Johan Bollen, Huina Mao, and Alberto Pepe. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. ICWSM, 11:450–453, 2011.
8. Eugene Y Bann and Joanna J Bryson. The conceptualisation of emotion qualia: Semantic clustering
9. J. Iannuzzi, K. Lundberg, and S. McKee. “The politics of socioeconomic status: how socioeconomic status may influence political attitudes and engagement.” Curr. Opin. Psychol., 18:11–14, 2017.
10. J. Levy Abitbol, et al. “Socioeconomic dependencies of linguistic patterns in twitter: A multivariate analysis”. In WWW TheWebConf’18, 1125-1134, 2018.
11. Hartmann, T. (2008). Parasocial interactions and paracommunication with new media characters. Mediated interpersonal communication, 177, 199.
12. Schramm, H., & Hartmann, T. (2008). The PSI-Process Scales. A new measure to assess the intensity and breadth of parasocial processes. Communications, 33(4), 385-401.
13. Loureiro, A. L., Miguéis, V. L., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems, 114, 81-93.
14. Zhang, S., Yao, L., Sun, A., & Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. ACM Computing Surveys (CSUR), 52(1), 1-38.

DOGUS

1. Shum, H. Y., He, X. D., & Li, D. (2018). From Eliza to XiaoIce: challenges and opportunities with social chatbots. Frontiers of Information Technology & Electronic Engineering, 19(1), 10-26.
2. Tunçer, M., Bilici, M. F., & Eryiğit, G. (2021, August). Development of Goal-Oriented Dialogue Systems for Customer Services in Automotive Industry. In 2021 International Conference on INnovations in Intelligent SysTems and Applications (INISTA) (pp. 1-7). IEEE.
3. Motger, Q., Franch, X., & Marco, J. (2022). Software-Based Dialogue Systems: Survey, Taxonomy and Challenges. ACM Computing Surveys (CSUR).
4. Adamopoulou, E., & Moussiades, L. (2020). Chatbots: History, technology, and applications. Machine Learning with Applications, 2, 100006.
5. Nivedita Bhirud, Subhash Tataale, Sayali Randive, and Shubham Nahar. 2019. A Literature Review On Chatbots In Healthcare Domain. International Journal of Scientiic & Technology Research 8, 7 (2019).
6. Lample, G., Ballesteros, M., Subramanian, S., Kawakami, K., & Dyer, C. (2016). Neural architectures for named entity recognition. arXiv preprint arXiv:1603.01360.
7. Nakov, P., Hoogeveen, D., Màrquez, L., Moschitti, A., Mubarak, H., Baldwin, T., & Verspoor, K. (2019). SemEval-2017 task 3: Community question answering. arXiv preprint arXiv:1912.00730.
8. Cer, D., Diab, M., Agirre, E., Lopez-Gazpio, I., & Specia, L. (2017). Semeval-2017 task 1: Semantic textual similarity-multilingual and cross-lingual focused evaluation. arXiv preprint arXiv:1708.00055.
9. Yang, Y., Yuan, S., Cer, D., Kong, S. Y., Constant, N., Pilar, P., ... & Kurzweil, R. (2018). Learning semantic textual similarity from conversations. arXiv preprint arXiv:1804.07754.
10. Reimers, N., & Gurevych, I. (2019). Sentence-bert: Sentence embeddings using siamese bert-networks. arXiv preprint arXiv:1908.10084.

EXPERTEAM

1. <https://www.appsruntheworld.com>

FTP-LDA & ISEP

1. Cheng, X., Bao, Y., Zarifis, A., Gong, W. and Mou, J. (2022), "Exploring consumers' response to text-based chatbots in e-commerce: the moderating role of task complexity and chatbot disclosure", [Internet Research](https://www.emerald.com/insight/publication/issn/1066-2243), Vol. 32 No. 2, pp. 496-517. <https://doi.org/10.1108/INTR-08-2020-0460>
2. OMD Optimal Management of Demand – project. https://itea4.org/community/project/omd/basics.html
3. Hussain, S., Ameri Sianaki, O., Ababneh, N. (2019). A Survey on Conversational Agents/Chatbots Classification and Design Techniques. In: Barolli, L., Takizawa, M., Xhafa, F., Enokido, T. (eds) Web, Artificial Intelligence and Network Applications. WAINA 2019. Advances in Intelligent Systems and Computing, vol 927. Springer, Cham. https://doi.org/10.1007/978-3-030-15035-8\_93
4. M.Mamatha, C.Sudha, L 2021, ’Chatbot for E-Commerce Assistance: based on RASA’, Turkish Journal of Computer and Mathematics Education Vol.12 No. 11 (2021), pp. 6173 – 6179,https://www.turcomat.org/index.php/turkbilmat/article/view/6943.
5. Xiaojie Wang, Caixia Yuan,Recent Advances on Human-Computer Dialogue,CAAI ansactions on Intelligence Technology,Volume 1, Issue 4,2016,Pages 303-312,ISSN 2468-2322,https://doi.org/10.1016/j.trit.2016.12.004.
6. Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow:

Concepts, tools, and techniques to build intelligent systems. O´Reilly Media.

1. Jurafsky, D., & Martin, J. H. (2020). Speech and language processing (3rd ed. Draft). Upper Saddle River, NJ: Prentice Hall.
2. Victoria Oguntosin, Ayobami Olomo, L 2021, “Development of an E-Commerce

Chatbot for a University Shopping Mall”, Applied Computational Intelligence and

Soft Computing, Article ID 6630326, 14 pages, 2021, <https://doi.org/10.1155/2021/6630326>

1. Sakshi Yadav, Raghvendra Pratap Singh, Dhanya Sree, Ms. A. Vidhya-vani, L 2021,” E-COMMERCE CHATBOT FOR PRICE NEGOTIATION”, vol 3, no. 11, https://www.irjmets.com/uploadedfiles/paper/volume 3/issue 11

november 2021/17105/final/fin irjmets1637170535.pdf

1. Lalwani, T., Bhalotia, S., Pal, A., Rathod, V., & Bisen, S. (2018). Implementation of a Chatbot System using AI and NLP. International Journal of Innovative Research in Computer Science & Technology (IJIRCST) Volume-6, Issue-3, <https://dx.doi.org/10.2139/ssrn.3531782>.
2. Open Sourcing BERT: State-of-the-Art Pre-training for Natural Language Processing. (2018b, November 2). Google AI Blog. Retrieved October 6, 2022, from<https://ai.googleblog.com/2018/11/open-sourcing-bert-state-of-art-pre.html>
3. spaCy 101: Everything you need to know · spaCy Usage Documentation. (n.d.). spaCy 101: Everything You Need to Know. Retrieved October 6, 2022, from<https://spacy.io/usage/spacy-101/>
4. Introduction to Rasa Open Source. (2022, October 6). Rasa Open Source Documentation. Retrieved October 6, 2022, from<https://rasa.com/docs/rasa/>
5. Why developers like UBIAI. (n.d.). StackShare. Retrieved October 6, 2022, from <https://stackshare.io/ubiai>
6. Mahmood, O. (2022, April 16). What’s Hugging Face? An AI community for sharing ML models and datasets | Towards Data Science. Medium. Retrieved October 6, 2022, from <https://towardsdatascience.com/whats-hugging-face-122f4e7eb11a>
7. Isinkaye, F. O., AbiodunBabs, I. G., & Paul, M. T. (2022). Development of a Mobile-Based Hostel Location and Recommendation Chatbot System. International Journal of Information Technology and Computer Science(IJITCS), Vol.14, No.3, pp.23-33, 2022. DOI: 10.5815/ijitcs.2022.03.03.
8. Santana, R., Ferreira, S., Rolim, V., de Miranda, P. B., Nascimento, A. C., & Mello, R. F. (2021). A Chatbot to Support Basic Students Questions. In LALA (pp. 58-67).
9. Illescas-Manzano, María D., Noé Vicente López, Nuno Afonso González, and Carmen Cristofol Rodríguez. 2021. "Implementation of Chatbot in Online Commerce, and Open Innovation" Journal of Open Innovation: Technology, Market, and Complexity 7, no. 2: 125. <https://doi.org/10.3390/joitmc7020125>
10. an Quy Tran, Thai Van Nguyen, Thang Duc Phung, Viet Tan Nguyen, Dat Duy Tran, and Son Tung Ngo. 2021. FU Covid-19 AI Agent built on Attention algorithm using a combination of Transformer, ALBERT model, and RASA framework. In 2021 10th International Conference on Software and Computer Applications (ICSCA 2021). Association for Computing Machinery, New York, NY, USA, 22–31. <https://doi.org/10.1145/3457784.3457788>