

THE BUCHAREST UNIVERSITY OF ECONOMIC STUDIES

DATABASE SYSTEMS JOURNAL

Vol. XIV/2023

LISTED IN

RePEc, EBSCO, DOAJ, Open J-Gate,
Cabell's Directories of Publishing Opportunities,
Index Copernicus, Google Scholar,
Directory of Science, Cite Factor,
Electronic Journals Library

BIG DATA

NoSQL

DATA SCIENCE

MACHINE LEARNING

BUSINESS INTELLIGENCE

CLOUD COMPUTING

DATA MINING

DATA WAREHOUSES

DATABASES

ISSN: 2069 – 3230
dbjournal.ro

Database Systems Journal BOARD

Director

Prof. Ion Lungu, PhD, University of Economic Studies, Bucharest, Romania

Editors-in-Chief

Prof. Adela Bara, PhD, University of Economic Studies, Bucharest, Romania

Conf. Iuliana Botha, PhD, University of Economic Studies, Bucharest, Romania

Editors

Conf. Anca Andreescu, PhD, University of Economic Studies, Bucharest, Romania

Conf. Anda Belciu, PhD, University of Economic Studies, Bucharest, Romania

Prof. Ramona Bologa, PhD, University of Economic Studies, Bucharest, Romania

Conf. Vlad Diaconița, PhD, University of Economic Studies, Bucharest, Romania

Lect. Alexandra Florea, PhD, University of Economic Studies, Bucharest, Romania

Prof. Adina Uța, PhD, University of Economic Studies, Bucharest, Romania

Editorial Board

Prof. Ioan Andone, PhD, A.I.Cuza University, Iasi, Romania

Prof. Emil Burtescu, PhD, University of Pitesti, Pitesti, Romania

Joshua Cooper, PhD, Hildebrand Technology Ltd., UK

Prof. Marian Dardala, PhD, University of Economic Studies, Bucharest, Romania

Prof. Dorel Dusmanescu, PhD, Petrol and Gas University, Ploiesti, Romania

Prof. Marin Fotache, PhD, A.I.Cuza University, Iasi, Romania

Dan Garlasu, PhD, Oracle Romania

Prof. Marius Guran, PhD, University Politehnica of Bucharest, Bucharest, Romania

Lect. Ticiano Costa Jordão, PhD-C, University of Pardubice, Pardubice, Czech Republic

Prof. Brijender Kahanwal, PhD, Galaxy Global Imperial Technical Campus, Ambala, India

Prof. Dimitri Konstantas, PhD, University of Geneva, Geneva, Switzerland

Prof. Hitesh Kumar Sharma, PhD, University of Petroleum and Energy Studies, India

Prof. Marinela Mircea, PhD, University of Economic Studies, Bucharest, Romania

Prof. Mihaela I.Muntean, PhD, West University, Timisoara, Romania

Prof. Stefan Nitchi, PhD, Babes-Bolyai University, Cluj-Napoca, Romania

Prof. Corina Paraschiv, PhD, University of Paris Descartes, Paris, France

Davian Popescu, PhD, Milan, Italy

Prof. Gheorghe Sabau, PhD, University of Economic Studies, Bucharest, Romania

Prof. Nazaraf Shah, PhD, Coventry University, Coventry, UK

Prof. Ion Smeureanu, PhD, University of Economic Studies, Bucharest, Romania

Prof. Traian Surcel, PhD, University of Economic Studies, Bucharest, Romania

Prof. Ilie Tamas, PhD, University of Economic Studies, Bucharest, Romania

Silviu Teodoru, PhD, Oracle Romania

Prof. Dumitru Todoroi, PhD, Academy of Economic Studies, Chisinau, Republic of Moldova

Prof. Manole Velicanu, PhD, University of Economic Studies, Bucharest, Romania

Prof. Robert Wrembel, PhD, University of Technology, Poznan, Poland

Contact

Calea Dorobanților, no. 15-17, room 2017, Bucharest, Romania

Web: <https://dbjournal.ro/>

E-mail: editordbjournal@gmail.com

CONTENTS

Prediction Model for Soybean Productivity	1
Ion GANEA	
Image search engine for furniture recommendations	16
Mariana ȘERBAN-PREDA	
Data Science Driven Analysis of the Impact of COVID-19.....	24
Șerban Gabriel TIMOFTE, Ioana Ruxandra STROE, Daniel TAȘCU, Raluca Cristina TERTEȘ, Radu Ioan MOGOȘ	
Internet of Things - Overview.....	33
Andreea MIHAI, Ștefania Codruța MĂNĂILĂ, Antonio Sebastian DUMITRAȘCU	
The Impact of Artificial Intelligence and the ChatGPT Tool on Society. A Case Study on Employees Working in Corporations in Romania	52
Elena COSTIN, Elena Daniela DUMITRAȘCU, Andreea Georgiana ENE, Florinela Georgiana PREDA	
Insights and proposals for RPA implementations	67
Andreea Izabela BOSTAN, Ana Ramona BOLOGA	

Prediction Model for Soybean Productivity

Ion GANEA

Moldova State University, Chisinau, Republic of Moldova
iganea9@gmail.com

Abstract. *This paper presents a holistic approach to biological and agricultural research focused on the use of interconnected technologies in the context of climate change. Researchers from different countries have analyzed how smart technologies can help agriculture adapt to these changes. The most representative works in the field are analyzed. Among these technologies are graph database systems such as Neo4j, which have demonstrated success in predicting the studied phenomena. The paper describes the development of a soybean crop productivity prediction model using monthly and annual data of meteorological phenomena such as precipitation, air temperature, hydrothermal coefficient, soil moisture and others. Some of the results of this promising research are also presented.*

Keywords: *Holistic , Knowledge, Models, Prediction, Graph, Neo4j, Graph Data Science.*

1 Introduction

Biological and agricultural systems are complex and can be better understood through a holistic approach. The concept of holism suggests that the components of organisms, biological networks, and the environment are more than their sum [1]. The term holistic comes from *holos* which means everything, and the holistic approach means looking at the problem from all points of view (or from above) [2]. This approach has led to significant developments in the life sciences and environmental sustainability, particularly by studying the interactions between the components of biological systems and how they describe the functions and behavior of the systems [3,4]. Systems biology is the computational and mathematical analysis and modeling of complex biological systems, integrating many scientific fields, including biology, computer science, bioinformatics, biophysics, bioengineering, and others [5]. This discipline aims to study the complex interactions within biological systems and uses a holistic approach to predict how these systems evolve and adapt to various economic, social and environmental conditions [6]. The goal is to develop durable, sustainable and resilient solutions to global problems such as food, health and the environment.

This mode of research refers to an integrated perspective that considers the interconnections and interdependencies between the various aspects of these fields. A certain phenomenon is determined by a number of factors. Each of the factors has an impact with a certain weight on the whole phenomenon. This approach aims to develop sustainable agricultural and economic systems that take into account environmental impact, plant health and productivity, human and animal health and welfare [7]. In addition, it provides a more comprehensive and accurate understanding of biological systems and helps to develop more sustainable and efficient practices in the management of natural resources[8]. Holistic research aims to promote more efficient and ecological agricultural practices that increase the sustainability of agricultural systems, ensuring adequate food production, while reducing the negative impact of agricultural activities on the environment and society. It promotes innovative agricultural practices, green technologies and conservation of natural resources.

2. Holistic Research in Biology and Agriculture.

In scientific research there are a lot of resources available, scientific papers that explore the holistic approach in biology, agri-

culture and agricultural economics and continue to develop as global problems become acute, emphasizing the need for better and more sustainable solutions to the problems in these fields. Among the most relevant works can be listed:

- "Ecological resilience, biodiversity, and scale" - published by Carl Folke and his collaborators in the *Annual Review of Ecology and Systematics*, this article explores the concept of ecological resilience and the importance of biodiversity in the development of sustainable systems. The authors emphasize the need for a holistic approach in addressing environmental problems, to avoid fragmented solutions limited to a single problem [9].
- "Organic Agriculture and the Global Food Supply" - published by Catherine Badgley and her colleagues in *Renewable Agriculture and Food Systems*, this article examines the potential of organic agriculture to meet global food demand sustainably. The authors argue that organic farming can help reduce environmental impact, improve human health and support the local agricultural economy [10].
- "The Holistic Management of Rangelands: A Model for Sustainable Livestock Production in Arid Zones" - published by Allan Savory and Jody Butterfield in the *Journal of Sustainable Agriculture*, this article describes a holistic method of managing rangelands in arid zones. The authors argue that this approach can support biodiversity and sustainable agricultural production, as well as improve the welfare of animals and local farming communities [11].
- "Agroecology: A Transdisciplinary, Participatory and Action-oriented Approach" by Miguel Altieri and Clara Nicholls. This article explores the concept of agroecology as a holistic approach to the development of sustainable agriculture, involving collaboration between researchers, farmers and local communities [12].
- "Holistic Management: A New Framework for Decision Making" by Allan Savory. This book presents the holistic approach to land management, which encourages decision-making based on the impact on ecosystems and local communities, rather than profit maximization [13].
- "Ecological Economics: Principles and Applications" by Herman Daly and Joshua Farley. This book presents the principles of ecological economics, which integrates economics with ecology and the limitations of natural resources and the environment in the process of economic development [14].
- "The Triple Bottom Line: How Today's Best-Run Companies Are Achieving Economic, Social, and Environmental Success - and How You Can Too" by Andrew Savitz. This book explores the concept of "triple impact" (economic, social and environmental) and how companies can integrate this approach into their business strategies [15].
- "Sustainable agriculture in Romania: the potential and challenges in the context of climate change" - published by Daniela Luca and her collaborators in the *Journal of Research and Social Intervention*, this article analyzes the potential and challenges of sustainable agriculture in Romania in the context of climate change. The authors emphasize the need for a holistic approach, which takes into account the complex interactions between the different components of the agricultural system and integrates sustainable agricultural practices, as well as technological and social innovations [16].

In the context of climate change, the holistic approach becomes increasingly important in the development of sustainable and resilient agricultural systems. This involves adopting an integrative perspective, which takes into account the complex interactions between the different components of the agricultural system, as well as its impact on the environment and local communities.

Here are some examples of relevant scientific papers in this field:

- "Agroecology and Sustainable Agriculture for Climate Change Adaptation and Mitigation: A Review" - published by Emmanuel Torquebiau and his collaborators in *Sustainability*, this article explores how agroecology can be used to address the challenges of climate change. The authors emphasize that agroecology can contribute to reducing greenhouse gas emissions, improving biodiversity and protecting natural resources, as well as increasing food security and the well-being of agricultural communities [17].
- "Climate Change and Agriculture: Advancing Knowledge and Mitigation Options" - published by Cynthia Rosenzweig and her colleagues in the *Annual Review of Environment and Resources*, this article analyzes the impact of climate change on agriculture and identifies options for adaptation and reduction of greenhouse gas emissions. The authors emphasize the need for a holistic approach, which takes into account the impact on soil, water and biodiversity, as well as agricultural communities and the local economy [18].
- "Climate Change Adaptation Strategies for Smallholder Farmers in the Sahel Region of West Africa: A Comprehensive Review" - published by Sifat Rehana and colleagues in *Sustainability*, this article examines climate change adaptation strategies for smallholder farmers in the Sahel region of West Africa. The authors emphasize that the holistic approach is essential to address the complexity of the environmental and social issues facing these farmers and to develop sustainable solutions [19].
- "Sustainable Agriculture and Climate Change: Producing Potatoes (*Solanum tuberosum* L.) and Bush Beans (*Phaseolus vulgaris* L.) for Improved Food Security and Resilience in a Canadian First Nations Community" - published by Mary Beckie and her collaborators in *Sustainability*, this article explores

how sustainable agriculture can contribute to improving food security and the resilience of local communities in the context of climate change. The authors emphasize the need for a holistic approach, which takes into account local needs and resources, as well as the complex interactions between the different components of the agricultural system [20].

- "Climate change and agriculture in the Republic of Moldova: opportunities and challenges" - published by Dumitru Ciorici and his collaborators in the *Moldovan Journal of the Environmental Science, Management and Engineering*, this article analyzes the impact of climate change on agriculture in the Republic of Moldova and identifies options for adaptation and reduction of greenhouse gas emissions. The authors emphasize the need for a holistic approach, which takes into account the complex interactions between the different components of the agricultural system and integrates aspects related to soil, water, energy and local communities [21].

3. Smart technologies in agriculture.

The use of smart technologies in agriculture is a relatively recent trend. In the context of global population growth, population aging in some parts of the world and labor shortages, the introduction of new technologies and the automation of agriculture are essential to solving problems in the field of agricultural development, the volume and quality of food products. In the agricultural field, intelligent technologies, such as Precision Agriculture (Agriculture 4.0) or Smart Agriculture (Agriculture 5.0), are innovative types of agriculture based on intelligent technologies, the implementation of digital tools, the Internet of Things (IoT) and working with data that aims to make the agricultural process more efficient [22]. Through these technologies, farmers can have an increased degree of control over the entire agricultural process, which improves the activity of farmers and authorities in the

field by providing adequate information for decision-making [23].

Graph databases can be used to develop machine learning models and algorithms that allow data analysis and prediction of biological and agricultural phenomena. These technologies can be used to identify complex relationships between various entities in an efficient manner. In biological research, graph systems are an industry standard for data modeling. They model different types of entities and relationships between them, using nodes and relationships between data to infer knowledge from existing relationships and to represent relationships from prior knowledge of the data [24].

4. Elaboration of the model for predicting the productivity of agricultural crops.

One of the common methods to predict the productivity of agricultural crops is the analysis of previous meteorological and agricultural data. This method is based on the fact that there is a correlation between weather conditions, such as temperature and humidity, and the productivity of agricultural crops. The models made can take into account several factors, such as soil quality, crop management, irrigation and fertilization levels, as well as weather conditions, to provide a more accurate prediction of productivity. Another method is the use of satellite images and sensor technology to monitor and evaluate agricultural crops. Researchers can use satellite images to collect information about plants, such as leaf colors and shapes, which can provide information about plant health and development. This data can then be processed through machine learning algorithms and used to estimate crop productivity.

In the field of agricultural crop productivity prediction, there are a number of relevant scientific works that have explored various methods and technologies to improve the ability to estimate agricultural production. Below are some examples of such works:

- "Agricultural yield prediction using machine learning: A review" by Chen et al. (2019) - This paper examines how machine learning technologies can be used to estimate agricultural crop productivity. The authors present a number of machine learning models that have been developed to predict the yield of cereals, vegetables and other crops, and discuss the advantages and disadvantages of these models [25].
- "Crop yield forecasting on a national scale using remote sensing data" by Thenkabail et al. (2019) - This paper describes a nationwide agricultural production prediction system that uses remote sensing data to estimate crop production. The authors present a series of data analysis algorithms that have been used to estimate the production of wheat, rice and other crops in various countries around the world [26].
- "Using machine learning techniques for crop yield prediction and climate change impact analysis" by Wang et al. (2020) - This paper presents an agricultural production prediction model based on machine learning technologies and meteorological data analysis. The authors discuss how this model can be used to estimate the production of maize and other crops, as well as to analyze the impact of climate change on agricultural production.
- "Evaluation of spectral indices for predicting grain yield of wheat using remote sensing data" by Fan et al. (2021) - This paper examines how remote sensing data can be used to estimate wheat production. The authors present a series of spectral indices that have been used to estimate wheat production in different geographic areas, and analyze the effectiveness of these indices in terms of prediction accuracy [28].

These are just a few examples of relevant work in the field of agricultural crop productivity prediction. However, researchers continue to explore and develop new methods and technologies to improve the ability to estimate agricultural production

and help farmers better manage their crops. In the Romanian space, a number of researchers have contributed to the field of predicting the productivity of agricultural crops through relevant scientific works. Below are some examples of such works:

- "Study on the factors influencing the productivity of wheat varieties grown under Moldovan conditions" by M. Tcaciuc and M. Pîrțac (2016) - This paper examines how different factors, such as fertilization, phytosanitary treatments and weather conditions, influence the productivity of the varieties of wheat grown in Moldova. The authors present a series of experimental data indicating that fertilization and phytosanitary treatments can have a significant impact on wheat production [29].
- "Using Remote Sensing in Agricultural Crop Productivity Prediction" by D. Chiriac and G. Dumitrașcu (2018) - This paper examines how remote sensing data can be used to estimate agricultural crop production in Romania and Moldova. The authors present a number of examples of the application of these technologies to estimate the production of corn, wheat and other crops, and discuss the advantages and disadvantages of using these technologies [30].
- "The WOFOST model for predicting the productivity of wheat crops under Moldovan conditions" by V. Chiriac and V. Popa (2018) - This paper presents a mathematical model to estimate wheat production in Moldova, based on meteorological data and soil characteristics. The authors demonstrate that this model can be successfully used to estimate wheat production and help farmers better manage their crops [31].

A powerful tool for the holistic approach to problem solving and sustainable development can be the Neo4j graph database system. This is a graph-based database management system that is used to model and store data in the form of nodes and relationships, making it a very effective sys-

tem to model and analyze complex relationships between data. This system can be used to support phenomenon prediction tasks. For example, relationships and data stored in a Neo4j database could be used as input to a machine learning model that makes predictions based on that data. In addition, Cypher, Neo4j's query language, can be used to identify patterns and trends in data, which could also be used as input to a prediction model. They allow the representation of complex relationships between different elements of a problem and their impact on the whole system.. These databases are used to represent and analyze data in a way that allows the identification and understanding of connections and interdependencies between different entities. These systems allow data to be represented in the form of graphs, which can be particularly useful for agricultural data analysis.

The use of graph database systems in the field of agricultural crop productivity prediction has been addressed by numerous researchers. Below are some examples of such works:

- "Using Graph Database for Soil Data Analysis and Agricultural Crop Production Prediction" by R. Wang et al. (2016) - This paper explores the use of a graph database to analyze soil data and predict agricultural crop production in an agricultural area in China. The authors demonstrate that the representation of data in the form of a graph can be very useful in identifying the complex relationships between factors that influence crop productivity [32].
- "Using Graph Databases for Weather Data Analysis and Agricultural Crop Production Prediction" by A. Rangwala et al. (2017) - This paper examines the use of a graph database to analyze weather data and predict agricultural crop production in an agricultural area in India. The authors present a series of examples that illustrate how graph analysis can be used to identify relationships between weather data and crop production [33].

- "Using Neural Networks and Graph Databases to Predict Corn Production in the United States" by M. Zhang et al. (2019) - This paper examines the use of neural networks and graph databases to predict corn production in the United States. The authors demonstrate that combining these technologies can lead to a significant improvement in prediction accuracy [34].
- "A Deep Learning Approach to Soybean Yield Prediction Using Field Observations and Satellite Imagery" by Ma et al. (2021). In this paper, the authors propose a new deep learning model to predict soybean crop yield using satellite imagery and field observation data [35].
- "Application of a Graph Database in Predicting Soybean Yield in Different Regions of Argentina" by H. de la Fuente et al. (2020). This paper presents a graph database model that is used to predict soybean crop yield in different regions of Argentina.
- "Graph Convolutional Networks for Soybean Yield Prediction: A Comparative Study" by Liu et al. (2020). In this paper, researchers compared the performance of graph-based convolutional neural network models to predict soybean crop yield in different regions of China [36].

All these works focus on using graph database systems to analyze data and make more accurate predictions regarding agricultural crop productivity. Graph database technology provides an efficient way to represent complex agricultural data and can be used to identify hidden patterns or relationships between variables. In the field of predicting the productivity of agricultural crops, various researches have been carried out using Neo4j graph database systems, among which the following scientific works can be distinguished:

- "Graph Database-Based Crop Yield Prediction Model Using Neo4j" by Liu et al. (2019). This paper proposes a method to build a graph database of agricultural data and use it to predict crop yield. Researchers have demonstrated the effectiveness of this approach by predicting

the yield of a maize crop in Shandong Province, China [37].

- "Design of a Knowledge Graph Based on Neo4j for Crop Growth and Yield Prediction" by Wang et al. (2020). This paper presents a method to construct a knowledge graph of agricultural crop growth and use this graph to predict crop yield. Researchers have demonstrated the effectiveness of this approach by predicting rice crop yield in Jiangsu Province, China [38].
- "A Graph Database Framework for Crop Yield Prediction" by Kumar et al. (2020). This paper presents a graph database framework for agricultural data analysis and used this framework to build a crop yield prediction model. Researchers have demonstrated the effectiveness of this framework by predicting potato crop yield in different regions of India [39].

These papers demonstrate the utility and effectiveness of using Neo4j graph database technology in the field of agriculture and crop productivity prediction. Neo4j enables more efficient analysis and visualization of complex agricultural data, which can lead to greater accuracy in crop yield prediction and better management of agricultural resources.

In order to obtain high and stable yields, it is necessary to study the cultivation of this crop through the use of intelligent technologies, including the prediction of productivity results and production quality. As a result of predicting productivity according to climatic conditions, important knowledge can be gained, including:

- Understanding the interaction between climate factors and productivity: By analyzing weather and production data, we can gain a deeper understanding of how weather conditions affect soybean productivity and identify patterns and trends in this relationship.
- Optimizing agricultural practices: With a better understanding of the interaction between weather conditions and soybean productivity, we can help farmers make

better decisions about agricultural practices, such as when to plant, type and amount of fertilizers, etc., to maximize soybean productivity and reduce costs and environmental impact.

- Improved production planning and supply chain management: With more accurate prediction of soybean productivity based on weather conditions, food companies can better plan the production and storage of soy-based food products, helping them reduce waste and costs and improve supply chain management.
- Improving sustainability: By identifying optimal agricultural practices to maximize soybean productivity depending on weather conditions, we can help reduce environmental impact and increase agricultural sustainability. Also, by improving production planning and supply chain management, we can reduce waste and costs and improve the efficiency and sustainability of the food industry.

In the specialized literature there are a number of scientific works carried out in different countries that use the Neo4j graph database systems in the field of predicting the productivity of soybean crops:

- "A Comparative Study of Neo4j and PostgreSQL in Soybean Yield Prediction" by Ye et al. (2021). This paper compares the performance of Neo4j and PostgreSQL graph databases in predicting soybean crop yield in different regions of China [40].
- "Graph-based Analysis of Soybean Yield Prediction Using Neo4j" by Hu et al. (2020). In this paper, researchers used Neo4j technology to build a graph database of field observation data and satellite images to predict soybean crop yield in different regions of China [41].
- "Development of a Soybean Yield Prediction System Using Neo4j Graph Database" by Wang et al. (2020). In this paper, the authors developed a soybean crop yield prediction system using the Neo4j graph database and evaluated the performance of the system in comparison with other prediction methods [42].

In the Romanian space, the use of these technologies is not reflected.

All of these works use Neo4j technology to build complex graph databases and make more accurate predictions of soybean crop productivity. Neo4j is a scalable and efficient graph database that can be used to analyze and visualize agricultural data. Agricultural research is constantly developing, and the use of graph database systems can be an interesting solution to analyze data and make more accurate predictions. Each research analyzes various aspects of certain problems specific to the area, the climatic conditions or the population in that area. Certainly, there are many research opportunities in this area. This is an example of the research potential in this area, where technology can be used to gain a better understanding of agricultural processes and make more accurate predictions of crop productivity.

The purpose of a model for predicting the productivity of agricultural crops, especially soy, is to assess the risks and to make the necessary decisions, to estimate the future production of the crop, to minimize the impact of climate change on the environment and human communities, to manage natural resources, as well as to develop sustainable economy.

The objectives of the productivity prediction model include:

1. Forecasting future production: One of the main objectives of an agricultural crop productivity prediction model is to estimate future crop production based on weather conditions and other variables.
2. Identifying influencing factors: The model can be used to identify influencing factors affecting soybean crop productivity and determine which of these factors have the greatest influence on production based on factors such as weather conditions, soil, type and amount of agricultural inputs (eg fertilisers, biostimulants, herbicides and pesticides) as well as other key variables

such as moisture level, temperature, light or pest infestation level.

3. Optimizing agricultural practices: The model can be used to identify best agricultural practices and optimize the use of agricultural inputs, as well as assist in real-time decision making during the cropping season.
4. Improved profitability: By providing accurate estimates of future production, the model can help increase profitability by reducing costs and increasing production.
5. Reducing risk and increasing resilience: The model can help reduce the risk associated with agricultural crops by identifying conditions that can affect productivity and providing timely information to enable farmers to take preventive measures.

Database development

To fulfill the objectives, a graph database was developed in Neo4j DBMS, Cypher language. Soybean productivity prediction using intelligent data analysis databases and graph models can be achieved through the following steps:

1. Collection and cleaning of relevant agricultural production data such as production history, meteorological data, soil type, use of plant products: fertilizers, pesticides, biostimulants as well as other important information.
2. Creating the database. Using the Cypher query language, nodes grouped according to data types (labels) are created and properties are added for the relevant variables;
3. Creating relationships between nodes to reflect interactions between factors that influence productivity to enable quick and easy access to needed information. For example, we can create a relationship between a node representing a rainfall amount and a node representing a biostimulator to show how a particular practice affects productivity.

Nodes were created to process data related to the amount of monthly and annual precipitation, monthly and annual average temperatures, monthly and annual hydrothermal coefficient, monthly and annual average wind speed, air quality (release of harmful substances into the atmosphere from factories and vehicles), the productivity in quintals per hectare (q/ha) of some species of agricultural plants (soy). Data from the period 2002 – 2022 published by the National Bureau of Statistics of the Republic of Moldova[43] are included.

The necessary nodes were created that contain the data related to the monthly and annual amount of precipitation, temperature, hydrothermal coefficient - *htc*), soybean productivity for the studied period. The data presented are from the central region of the Republic of Moldova.

```
MERGE (jan2rC:MYRC{name:'Jan 2002 Rainfall Center', Rainfall:17}),
(feb4rC:MYRC{name:'Feb 2004 Rainfall Center', Rainfall:38}),
(may15rC:MYRC{name:'May 2015 Rainfall Center', Rainfall:15});
MERGE (mar3tC:MYTC{name:'Mar 2003 Temperature Center', temp:7.2}),
(apr10tC:MYTC{name:'Apr 2010 Temperature Center', temp:11});
MERGE (apr2chtC:MYHTC{name:'Apr 2002 HTC Center', htc:1.0, drought:0}),
(jul6chtC:MYHTC{name:'Jul 2006 HTC Center', htc:0.8, drought:1});
MERGE (r2c:YRC{name:'Rainfall 2002 Center', rainfall:604}),
(r6c:YRC{name:'Rainfall 2006 Center', rainfall:564});
MERGE (t5c:YTC{name:'Temperature 2005 Center', temp:10.5}),
(t8c:YTC{name:'Temperature 2008 Center', temp:11.3});
MERGE (y5C:YCHTC{name:'CHT 2005 Center', cht:1, drought:0}),
(y9C:YCHTC{name:'CHT 2009Center', cht:0.8, drought:2});
```

Nodes were made regarding soybean productivity for the studied period:

```
MERGE (soy05:Productivity{name:'Soy Prod 2005', harvest:18}),
(soy11:Productivity{name:'Soy Prod 2011', harvest:14});
```

A fragment from the graph database is represented in figure 1.

through machine learning methods can be achieved by creating graph processing pipelines (*Graph Processing Pipelines*) which represent a set of processes and technologies used to perform analysis and processing operations on graphs, i.e. on data models that consist of interconnected nodes and relationships. In the GDS library in the Neo4j DBMS, pipelines provide an end-to-end workflow from feature extraction to training and applying machine learning models. These pipelines are commonly used in data analytics and machine learning to perform a number of operations such as:

- Data extraction from nodes and relationships
- Graph metrics and properties such as graph centrality and node degree computation
- Graph analysis and modeling such as clustering and community detection
- Machine learning models training and testing to perform classifications and predictions on graphs.

Creating a prediction pipeline:

The first step in developing a crop estimation model is to design an in-memory

graph, which can later be used in algorithms and graph queries.

The following statement will project a graph using a native projection using the function `"gds.graph.project"`, which can later be used in graph algorithms and queries, and store it in the graph catalogue under the name `"predProd"`, using the tags (types of nodes): `":YHTCc"` (the nodes containing the annual hydrothermal coefficient - *htc*, with the properties `"htc"` and `"drought"`, based on which the production quantity can be determined), `":Prod"` (with the properties `"harvest"` and `"class"`, for training) and `":ProdP"` (with the property `"harvest"`, for prediction - the average harvest expressed in quintals per hectare - q/ha). The result is shown in figure 2.

```
CALL gds . graph . project ( 'prod-
Pred' ,
  { YHTCc : { properties : [ 'htc' ,
'drought' ] } ,
  Prod : { properties : 'harvest' } ,
  ProdP : { properties : 'harvest' }
} ,
  { DETERMINE_PROD : {
    type : 'DETERMINE_PROD' ,
    orientation : 'UNDIRECTED'
  } } ) ;
```

nodeProjection	relationshipProjection	graphName	nodeCount	relationshipCount
{Prod: {label: "Prod", properties: {harvest: {defaultValue: null, property: "harvest"}}, YHTCc: {label: "YHTCc", properties: {drought: {defaultValue: null, property: "drought"}, htc: {defaultValue: null, property: "htc"}}}}	{DETERMINE_PROD: {orientation: "UNDIRECTED", indexInvers: false, aggregation: "DEFAULT", type: "DETERMINE_PROD", properties: {}}}	"prodPred"	37	34

Fig. 1. The projected graph.

In order to be able to analyze the complex interactions between the different elements and variables that can affect the appearance and evolution of the studied phenomena, the incorporation of nodes into a model (embedding) is used as a pre-processing step. Nodes represent the interconnected objects or entities in a given context, thus

reflecting the relationships and interdependencies between them. Node embedding algorithms compute reduced-dimensional vector representations of the nodes in a graph. These vectors, also called embeddings, can be used for machine learning and are typically used as input for machine learning tasks.

```
CALL gds.fastRP.write('prodPred',{ embeddingDimension: 20,
writeProperty: 'fastrp-embedding'})
YIELD nodePropertiesWritten;
```

A common technique used in machine learning and data analysis is to determine similar nodes as part of preprocessing using *kNN* (*k-Nearest Neighbors*). In this technique, the *kNN algorithm* is used to find the *k* nearest neighbours of a given node in a data set based on a similarity metric. Once the *k* nearest neighbors are identified, they can be used to group similar nodes together for clustering or classification tasks, or they could be used to identify outliers or anomalies in the data.

```
CALL gds.knn.write('dtPred', {
writeRelationshipType: 'SIMILAR',
writeProperty: 'score', topK: 1, similarityCutoff: 0.8,
randomSeed: 42, concurrency: 1,
nodeProperties: ['harvest']})
YIELD nodesCompared, relationshipsWritten
```

The *gds.knn.write* function returns the number of nodes compared and the number of relationships created in the graph, which are saved in the *nodesCompared* and *relationshipsWritten* variables, respectively.

The *gds.knn.write* function receives the following arguments:

- *'prodPred'*: the name of the graph in which the similarity relations are to be created
- *writeRelationshipType*: the type of relationship that will be created between similar nodes (in this case '*SIMILAR*')
- *writeProperty*: the property to be added to the relation (in this case 'score'). The similarity score is a numerical value that measures the degree of similarity between two nodes. In machine learning and information retrieval, the similarity score is used to rank search results based on their relevance to a query. This is often done by comparing the query to each entity in a database and calculating a similarity score for each. The entities with the highest similarity scores are then returned as the top search results.

- *topK*: the maximum number of similar nodes that will be connected by the 'SIMILAR' relationship for each node (in this case, 1)
 - *similarityCutoff*: the minimum similarity threshold that must be reached for two nodes to be considered similar and therefore connected by the 'SIMILAR' relationship (in this case, 0.5)
 - *randomSeed*: the value used to initialize the random number generator, to ensure the reproducibility of the experiment
 - *concurrency*: the maximum number of threads that can be used simultaneously to process data
 - *nodeProperties*: list of node properties to consider in the similarity calculation.
- a) Creating a pipeline named " *pipe* " for training:

```
CALL gds.beta.pipeline.nodeClassification.create('pipe')
```

- b) Pipeline configuration. Adds a node property step to the pipeline. Here the input graph contains a *harvest node property*:

```
CALL gds . beta . pipeline . nodeClassification . addNodeProperty (
'pipe' , 'alpha.scaleProperties' , {
nodeProperties : [ 'harvest' , 'class'
] ,
scaler : 'L1Norm' , mutateProperty :
'scaledSizes' })
YIELDS name , nodePropertySteps ;
```

- c) Selecting Pipe Features and Pipe Splitting:

```
CALL gds.beta.pipeline.nodeClassification.selectFeatures(
'pipes', ['scaledSizes', 'harvest',
'class'])
YIELDS name, featureProperties;
```

Training the pipeline

The training data is used to train the prediction model. This can be achieved by using different machine learning techniques such as linear regression, neural networks or classification algorithms. The resulting table is represented in figure 3.

```
CALL
gds.beta.pipeline.nodeClassification.train('prodPred', {
```



```

pipeline: 'pipe', targetNodeLabels:
['Prod'],
modelName: 'nc-pipeline-model',
targetProperty: 'harvest',
randomSeed: 1337, metrics: ['ACCURACY',
'OUT_OF_BAG_ERROR']
}) YIELD modelInfo, modelSelectionStats
RETURN
modelInfo.bestParameters AS
winningModel,
modelInfo.metrics.ACCURACY.train.avg
AS avgTrainScore,
modelInfo.metrics.ACCURACY.outerTrain
AS outerTrainScore,
modelInfo.metrics.ACCURACY.test AS
testScore,
[cand
IN
modelSelectionStats.modelCandidates|ca
nd.metrics.ACCURACY.validation.avg] AS
validationScores;

```

A fragment of the resulting of pipeline training is shown in figure 3.

winningModel	avgTrainScore	outerTrainScore	testScore	validationScores
{maxEpochs: 100, minEpochs: 1, cla ssWeights: [], penalty: 0.0, patie nce: 1, methodName: "LogisticRegre ssion", focusWeight: 0.0, batchSiz e: 100, tolerance: 0.001, learning Rate: 0.001}	0.6927272739999999	0.15384616	0.0	[0.700000002, 0.16666666800000002, 0.0, 0.700000002, 0.700000002]

Fig. 2. The result of pipeline training.

Model Validation: After the model has been trained, it must be validated to ensure that it is working correctly. This is done by using the test data set to evaluate the performance of the model. For example, you can use metrics such as accuracy, sensitivity, and specificity to evaluate model performance.

Testing the prediction model using historical data and adjusting it if necessary to improve its accuracy. Result prediction is representing in the figure 4.

```

CALL
gds.beta.pipeline.nodeClassification.p
redict.stream('prodPred', {
modelName: 'nc-pipeline-model',
includePredictedProbabilities: true,
targetNodeLabels: ['ProdP']
})
YIELD nodeId, predictedClass,
predictedProbabilities
WITH gds.util.asNode(nodeId) AS
harvestNode, predictedClass,
predictedProbabilities
RETURN harvestNode.name AS
classifiedYear, predictedClass,
floor(predictedProbabilities[predicted
Class] * 100) AS confidence
ORDER BY classifiedYear

```

classifiedYear	predictedClass	confidence
"Soy Prod 2019 Center"	14	95
"Soy Prod 2020 Center"	14	96
"Soy Prod 2021 Center"	20	95

Fig. 3. Prediction result.

Using the model: After the model has been validated, it can be used to make predictions for new data or to explore data in the database.

5. Conclusions

The holistic approach to crop cultivation under climate change involves the use of a set of agricultural measures and practices that improve crop performance, protect soil and water, and ensure sustainable and quality production. These measures must be adapted to the specific needs and conditions of each agricultural holding, taking into account the impact of different factors and the weight of each one in the final result. The use of intelligent technologies can significantly contribute to solving structured, unstructured and poorly structured problems in agriculture under the conditions of climate change. These technologies can improve the efficiency and accuracy of decision-making, help identify the best agricultural practices, and increase the productivity and sustainability of

agriculture. In addition, the initial costs of implementing new technologies can be high, which can make their adoption unaffordable for small and medium-sized farmers. Also, the use of these technologies may require continuous investment over time, which may be difficult for farmers facing limited returns. Efforts to adapt to new smart technologies can help improve agricultural performance, increase profitability and ensure the sustainability of the agricultural sector.

Prediction is an ever-evolving field, and researchers continue to explore and develop new methods and technologies to improve crop productivity prediction. These methods include the analysis of past weather and agricultural data, the use of satellite imagery and sensor technology, as well as data analytics and artificial intelligence technologies. The use of intelligent data analysis databases and graph models can be a valuable tool to help farmers and agricultural entrepreneurs optimize their agricultural production and make informed and sustainable decisions. The created model uses the data from the period 2002 – 2018. For testing, the data from the years 2019 – 2021 was checked. These predictions have mostly come true.

The created model can also be used for various agricultural crops, but also for predicting the evolution of climatic factors and increasing resilience to various natural hazards.

Acknowledgments

I would like to thank the Associate professor, PhD. Ana Bârsan, Head of the Department of Biology and Ecology, Faculty of Biology and Geoscience, Moldova State University for the data obtained during the experiments and the valuable consultations in the field of plant physiology, and professor, Gheorghe Capatână for methodological assistance related to research and development of the intelligent support system.

Notes: The work was carried out as part of the Doctoral Project "Models, Techniques and Program Products for

Intelligent Data Analysis in Plant Physiology". Scientific advisor: PhD. Eng. Gheorghe Capatână, MSU.

References

- [1] JC Smuts. *Holism and Evolution*. Macmillan, 1926.
- [2] Scribd <https://ro.scribd.com/document/351967625/Abordarea-Holistica#>. (accessed on 2023-03-01)
- [3] A. Savory, "Holistic Management: A New Framework for Decision Making," Island Press, 1999.
- [4] DEX <https://dexonline.ro/definitie/holism>. (accessed on 2023-03-04)
- [5] JJ Tyner and RE Michod, "Exploring the Complexity of Biological Systems," *BioScience*, vol. 53, no. 10, pp. 976-981, Oct. 2003.
- [6] H. Kitano, "Systems Biology: A Brief Overview," *Science*, vol. 295, no. 5560, pp. 1662-1664, Mar. 2002.
- [7] H. Daly and J. Farley, "Ecological Economics: Principles and Applications," Island Press, 2004.
- [8] HMI. Christine Jones. *The Back Forty Down Under: Adapting Farming to Climate Variability*. Available: <https://holisticmanagement.org/wp-content/uploads/2016/08/Jones-AdaptingFarming-Quivira-Feb10.pdf>. (accessed on 2023-02-24)
- [9] C. Folke et al., "Ecological resilience, biodiversity, and scale," *Annual Review of Ecology and Systematics*, vol. 31, pp. 277-308, 2000.
- [10] Badgley et al., "Organic Agriculture and the Global Food Supply," Cambridge University Press, 2007.
- [11] Savory and J. Butterfield, "The Holistic Management of Rangelands: A Model for Sustainable Livestock Production in Arid Zones," *Journal of Sustainable Agriculture*, vol. 4, no. 2-3, pp. 141-154, 1994.
- [12] M. Altieri and C. Nicholls, "Agroecology: A Transdisciplinary, Participatory and Action-oriented Approach," CRC Press, 2017.
- [14] A. Savory, "Holistic Management: A New Framework for Decision Making," Island Press, 2016.

- [13] H. Daly and J. Farley, "Ecological Economics: Principles and Applications," Island Press, 2010.
- [14] A. Savitz, "The Triple Bottom Line: How Today's Best-Run Companies Are Achieving Economic, Social, and Environmental Success - and How You Can Too," Jossey-Bass, 2006.
- [15] D. Luca et al., "Sustainable agriculture in Romania: potential and challenges in the context of climate change," *Revista de Research si Interventie Socială*, vol. 60, pp. 7-21, 2018.
- [16] E. Torquebiau et al., "Agroecology and Sustainable Agriculture for Climate Change Adaptation and Mitigation: A Review," *Sustainability*, vol. 8, no. 6, pp. 1-23, 2016.
- [17] C. Rosenzweig et al., "Climate Change and Agriculture: Advancing Knowledge and Mitigation Options," *Annual Review of Environment and Resources*, vol. 37, pp. 1-32, 2012.
- [18] S. Rehana et al., "Climate Change Adaptation Strategies for Smallholder Farmers in the Sahel Region of West Africa: A Comprehensive Review," *Sustainability*, vol. 11, no. 13, pp. 1-31, 2019.
- [19] M. Beckie et al., "Sustainable Agriculture and Climate Change: Producing Potatoes (*Solanum tuberosum* L.) and Bush Beans (*Phaseolus vulgaris* L.) for Improved Food Security and Resilience in a Canadian First Nations Community," *Sustainability*, vol. 7, no. 9, pp. 1-18, 2015.
- [20] D. Ciorici et al., "Climate change and agriculture in the Republic of Moldova: opportunities and challenges", *Moldovan Journal of the Environmental Science*, vol. 12, no. 2, pp. 50-60, 2020.
- [21] E. Costa, M. Martins, E. Vendruscolo, A. Silva, T. Zoz, F. Binotti, T. Witt, and C. Seron, "Greenhouses within the Agriculture 4.0 interface," *Rev. Ciencia Agronomica*, vol. 51, article e20207703, 2021. DOI: 10.5935/1806-6690.20200089.
- [22] Jin, B. Chen, BD Lampinen, and PH Brown, "Advancing Agricultural Production with Machinetesting Analytics: Field Determinants for California Almond Orchards," in *Proc. IEEE Int. Conf. on Agricultural and Biological Systems Engineering (ABSE)*, 202X, pp. 1-6.
- [23] Neo4j. Graph Data Modeling. <https://neo4j.com/graphacademy/edu/program/>. (accessed on 2023-03-30)
- [24] S. Chen, Y. Chen, X. Yang, and J. Zhang, "Agricultural yield prediction using machine learning: A review," *Computers and Electronics in Agriculture*, vol. 163, 2019, p. 104859. <https://www.sciencedirect.com/science/article/pii/S0168169918306782>
- [25] PS Thenkabail, MK Gumma, P. Teluguntla, A. Oliphant, J. Xiong, C. Giri, V. Dheeravath. Crop yield forecasting on a national scale using remote sensing data. *Remote Sensing*, 11(23), 2019, p. 2821.
- [26] B. Wang, X. Tao, H. Wang, J. Liu, Y. Zhang. Using machine learning techniques for crop yield prediction and climate change impact analysis. *Environmental Research*, 188, 2020, 109825.
- [27] X. Fan, Y. Shen, H. Zhang, L. Sun. Evaluation of spectral indices for predicting grain yield of wheat using remote sensing data. *International Journal of Remote Sensing*, 42(8). 2021, 2742-2759.
- [28] M. Tcaciuc, M. Pîrțac. Study on the factors that influence the productivity of wheat varieties grown in the conditions of Moldova. *Bulletin of the Academy of Sciences of Moldova. Mathematics*, 1(76), 3-12, 2016.
- [29] D. Chiriac, G. Dumitrașcu. The use of remote sensing in the prediction of agricultural crop productivity. *Studia Universitatis Moldaviae. Life Sciences Series*, 11(1), 2018, 89-93.
- [30] V. Chiriac and V. Popa, "The WOFOST model for predicting the productivity of wheat crops in the conditions of Moldova," *Rev. Agric. Econ.*, vol. 68(2), pp. 51-58, 2021.
- [31] R. Wang et al., "Using graph database for soil data analysis and crop yield prediction," in *Proceedings of the*

- 2016 7th International Conference on Intelligent Systems, Modeling and Simulation (ISMS), Bangkok, Thailand, 2016, pp. 42-46.
- [32] A. Rangwala et al., "Using graph databases for weather data analysis and crop yield prediction," in Proceedings of the 2017 8th International Conference on Intelligent Systems, Modeling and Simulation (ISMS), Kuala Lumpur, Malaysia, 2017, pp. 82-86.
- [33] M. Zhang et al., "Using neural networks and graph databases to predict corn production in the United States," *Rev. Agric. Econ.*, vol. 66(2), pp. 35-41, 2019.
- [34] W. Ma et al., "A Deep Learning Approach to Soybean Yield Prediction Using Field Observations and Satellite Imagery," *Remote Sens.*, vol. 13, no. 2, pp. 354, Jan. 2021.
- [35] H. de la Fuente et al., "Application of a Graph Database in Predicting Soybean Yield in Different Regions of Argentina," *Sensors*, vol. 20, no. 22, p. 6543, 2020.
- [36] H. Liu et al., "Graph Convolutional Networks for Soybean Yield Prediction: A Comparative Study," *IEEE Access*, vol. 8, pp. 81794-81805, 2020.
- [37] R. Wang et al., "Using graph database for soil data analysis and crop yield prediction," Proceedings of the International Conference on Computers Communications and Control (ICCCC), pp. 527-532, 2016.
- [38] R. Kumar, S. Bhandary, A. Patil, and M. Sonawane, "A Graph Database Framework for Crop Yield Prediction," in 2020 IEEE 10th International Conference on Intelligent Systems (IS), 2020, pp. 516-521. <https://ieeexplore.ieee.org/document/9263211>.
- [39] Y. Ye, J. Yang, X. Zhang, J. Wang, and Y. Gao, "A Comparative Study of Neo4j and PostgreSQL in Soybean Yield Prediction," in 2021 International Conference on Artificial Intelligence and Big Data Engineering (ICAIBE), 2021, pp. 260-265. <https://ieeexplore.ieee.org/document/9441644>
- [40] Y. Hu, S. Wang, X. Zhu, and C. Li, "Graph-based Analysis of Soybean Yield Prediction Using Neo4j," in 2020 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC), 2020, pp. 449-455. <https://ieeexplore.ieee.org/document/9259055>. (accessed on 2023-03-24)
- [41] Y. Wang, X. Zhao, and G. Zhang, "Development of a Soybean Yield Prediction System Using Neo4j Graph Database," in 2020 4th International Conference on Intelligent Information Processing (ICIIP), 2020, pp. 1-5. <https://ieeexplore.ieee.org/document/9277286>
- [42] Moldova Statistical Data Bank. <https://statbank.statistica.md/PxWeb/pxweb/ro/? rxdid= 2f5a0079-c3c1-461e-81a4-d854b7ba0fcd>. (accessed on 2023-02-24)
- [43] State Hydrometeorological Service. <http://www.meteo.md/index.php/clima/terminologie-si-unitati-de-masura/glosar-de-termeni/c/> (accessed on 2023-02-25).



Ion GANEA, master in Exact Sciences (Databases & Knowledge). PhD student at Doctoral School of Physical, Mathematical, Information and Engineering Sciences, Moldova State University. Teacher of Information Technology at the Center of Excellence in Economics and Finance, Computer Science Department. Courses taught: Databases; Computer Networks; Operating Systems. Areas of scientific interest: Artificial Intelligence, Intelligent Systems, Graph Databases, Intelligent Data Analysis.

Image search engine for furniture recommendations

Mariana ȘERBAN-PREDA
The University of Economic Studies, Romania
serbanpredamariana18@stud.ase.ro

Artificial intelligence (AI) has made enormous strides in recent years, transitioning from science fiction to a technology that is revolutionizing every sector of the global economy. Thanks to advances in machine learning, natural language processing and computer vision, artificial intelligence is no longer a futuristic dream, but a reality today. Once optimised and integrated into everyday life, artificial intelligence will substantially enhance human capabilities and contribute to the betterment of society. This paper will embrace the opportunities offered by computer vision, as part of artificial intelligence, by showcasing the FurnishMe application, an image search engine for furniture recommendations. Buying furniture online, as well as offline, is an overwhelming process given the quantity and diversity of furniture products available. The FurnishMe software solution allows users to easily explore the furniture of the three largest furniture retailers in Romania: Ikea, Jysk, and Dedeman. The system analyses user-uploaded interior design images in order to identify furniture items and provide aesthetically similar products from the three big retailers mentioned above. Both consumers and traders benefit from this solution. Clients benefit from a quick and easy way to choose the product they desire which integrates various features such as design, texture and colour. Moreover, businesses gain from greater sales by luring clients and saving time on in-person consultations.

Keywords: Deep Learning, Convolutional Neural Networks, Object Detection, Furniture recommendations, Web Scraping, Computer Vision, Visual Search Engine

1 Introduction

In recent years, the technology innovation has helped propel e-commerce forward, making online purchases of furniture increasingly commonplace. As civilization has evolved and living standards have increased, people have inclined to establish visual comfort in their houses in addition to safety and physical comfort. As part of an increasingly digital environment, it's simply natural for technology developers to search for ways of providing an easy and personalized shopping experience when purchasing furniture online. According to Ozturkan [1], 8 out of 10 consumers who have been impressed by giants like IKEA or Amazon have bought furniture online thanks to hyper-personalised marketing, augmented reality and the ability to contextualise potential purchases in their own home.

E-commerce and the furniture industry have experienced impressive growth over recent years. However, searching for and choosing

furniture online can be a challenging and time-consuming activity which decreases user satisfaction through poor choices or limited options available.

Furniture shopping can be an overwhelming task for the majority due to the variety of furniture products on the market. Moreover, searching through specialized website catalogues is a time-consuming activity since there are no compatibility checks between textures, colours or dimensions of the products. Consumers are frequently inspired by interior design images on social media, yet they may find it challenging to discover specific or similar furniture products. This step requires browsing through multiple online stores for products with matching features.

The aim of this study is to create a computer vision technology-based information system that simplifies the home furnishing process. Based on the recognition of furniture objects, the system will generate relevant

and personalised product recommendations from partner furniture stores. As a result, users will benefit from an improved experience in finding the right furniture for their needs and preferences.

In the next sections of this paper, we will present methods, technologies and results obtained during this research project in order to demonstrate its efficiency and usefulness for furniture recognition and recommendation.

2. Literature Review

The information technology system to be implemented is an image-based search engine that stands out from other similar systems by integrating a furniture object detection module.

Today, computer vision-based information systems provide deep and truthful analysis of visual reality thanks to deep learning techniques and convolutional neural networks (CNNs). These techniques have been applied in various fields such as fashion, medicine, transport or furniture. In the furniture industry, image-based recommendation engines can identify similar products based on design, style and various characteristics such as colour or texture.

It is essential to understand the accuracy and applicability of these techniques and to highlight new research directions.

In the paper “*DeepStyle: Multimodal Search Engine for Fashion and Interior Design*”, Trzcinski and his collaborators developed a multimodal search engine that combines visual and textual cues to retrieve elements aesthetically similar to the query from a multimedia database [2].

They aim to enable intuitive retrieval of products for fashion and interior design domains, as well as to address the inadequacies of currently existing search engines, by using a neural network architecture to model the joint multimodal space of database objects.

The search engine accepts two types of query information: an image that contains object(s) and a textual query that specifies

the search criteria. YOLO detection model was used in order to detect objects of interest. Visual search identifies the image query's nearest neighbours in the space of extracted features. Textual queries limit the set of results to those that are relevant to the query.

The article discusses several neural network architectures used for multimodal search, including convolutional neural networks (CNNs), Siamese Networks, LSTM networks, and structure-content neural language models (SC-NLMs). The authors propose their own architecture called *DeepStyle* which uses a combination of these elements [2].

Another relevant study to this research presents a new interior style detection approach that uses multi-scale features and boosting to increase accuracy over traditional and residual network methods. With this approach, interior style detection involves non-hierarchical clustering and multi-scale feature fusion by using spatial pyramid matching (SPM), colour information and object detection into bag of visual words (BoVW) [3]. Features extracted from images may include colour histogram, vector colour analysis and local feature histograms. The algorithm also involves predetermining cluster number in advance, training with rule-based boosting and using LightGBM to estimate room style. The proposed method outperforms conventional BoVW methods and residual network (ResNet) in terms of accuracy.

The literature has highlighted various approaches to the development of image-based search engines. Considering the detailed notions previously discussed, the primary objective of this paper is to integrate advanced image processing algorithms into an efficient and relevant solution for the needs of the furniture industry and consumers.

p_c	b_x	b_y	b_w	b_h	c_1	c_2	c_n
-------	-------	-------	-------	-------	-------	-------	-----	-----	-------

3. Methodology

The architecture of the proposed solution is composed of several interconnected components that work together to provide functionalities like object detection and furniture recommendation. Figure 1 shows a diagram of the solution architecture:

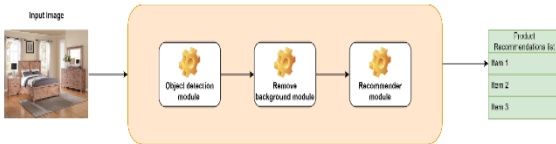


Figure 1. Solution architecture

A. Object Detection Module

YOLOv5 neural network was trained to detect pieces of furniture. You Only Look Once (YOLO) is a state-of-the-art algorithm for real-time object detection. Outstanding features for this algorithm are speed and accuracy in detection. YOLO uses a single neural network to perform classification as well as calculating the coordinates of the bounding boxes of objects in an image. It is currently an open-source algorithm that benefits from the support of a large community that constantly contributes to its improvement [4].

The first step in object detection using the YOLO algorithm starts by dividing the original image into $N \times N$ cells of equal shapes. Thus, each cell is responsible for localizing and classifying the object that it covers, along with a confidence score. This confidence score reflects the model's certainty that an object exists in that cell and that the delineation is accurate [4,5].

The next step is to determine the bounding boxes. YOLO determines the values for the bounding boxes' attributes using a regression algorithm where Y is the final vector representation for each bounding box and has the following structure [4]:

- p_c represents the probability score of the cell containing an object;
- (b_x, b_y) indicates the centre of the predicted bounding box;

- (b_w, b_h) represents the size of the bounding box;

- c_i corresponds to the class of objects predicted by the classification algorithm

Frequently, there may be multiple bounding box predictions for a single object. Using the Intersection over Union (IoU) and Non-Max Suppression metrics we manage to eliminate redundant detections. IoU measures the overlap between two bounding boxes which are usually the predicted box and the ground truth box. The metric has a range from 0 to 1 and is defined as the ratio of the area of intersection to the area of meeting of two rectangles as shown in Figure 2. The higher the value of IoU, the more the predicted box overlaps with the ground truth box. To eliminate redundant boxes, the NMS metric compares the previously calculated scores and eliminates those boxes with lower scores to the advantage of overlapping boxes with a higher score [6].

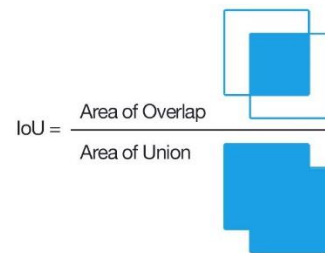


Figure 2. Intersection over union

Architecturally, the YOLO model consists of three key components: head, neck and backbone as shown in Figure 3. The backbone component is composed of convolutional layers. It is responsible for extracting features from an input image. Firstly, it is trained on a classification dataset, such as ImageNet. Also, the backbone component is usually trained with a lower resolution than the final detection model, since detection requires finer details than classification [5]. The neck component uses the features of convolution layers in the fully connected layered backbone to make predictions about the probabilities and bounding box coordinates. The head component is the final output layer of the network that can be interleaved with other

layers with the same input shape for transfer learning. These three portions of the model work together for feature extraction, object classification and localization [5,7].

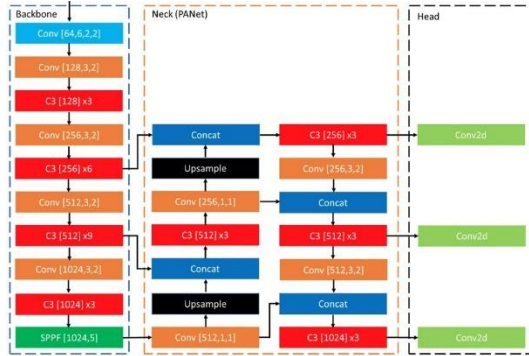


Figure 3. YOLO architecture [7]

B. Background removal

Background removal is a key technique which improves the recommendations of furniture products. The only relevant features for finding similar objects are the features of the detected object and not the environment. Consequently, we used Mask R-CNN model to segment the detected object and to translate it into an image with a white background. Thus, by reducing the background noise we manage to improve the accuracy of the recommendation algorithm. Due to the lack of a dataset including the object masks of the FurnishMe application's eight classes, we used the pre-trained model, provided by the open-source platform Detectron2, whose modules could be used by simply installing them in the Anaconda virtual environment.

C. Integrating with Ikea, Jysk, Dedeman store websites

The web scraping process allows the application to be integrated with the most popular e-commerce sites in the furniture industry. The Selenium library was used to browse the websites of IKEA, Jysk, and Dedeman stores, while BeautifulSoup, a library for parsing HTML and XML documents, was used to extract data. Images of products linked to the eight categories were downloaded and saved to disk, and details such as name, price, and external link were saved in the database.

D. Reverse Image Search Engine

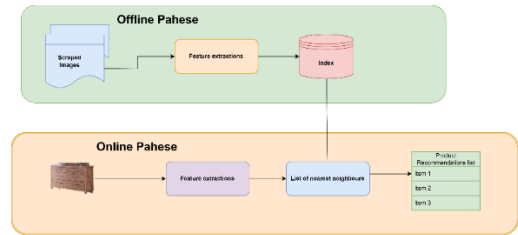


Figure 4. Image Search Engine architecture

In terms of recommending furniture products from the websites of Romania's most popular retailers, our primary goal was to develop a module that relies solely on visual similarity. It was identified the need to transform the information contained in millions of pixels of an image into a small representation that can be compared with other representations. A convolutional neural network takes an image as input and converts it into a feature vector. The latter serves as input to a classification layer which identifies the class label [2]. An ideal way to find similar images would be to pass the images through a high-performance convolutional neural network, extract the features and use a metric such as Euclidean distance or angular cosine to calculate the closest feature vectors.

Resnet neural network was used to understand the distinctive features of the images and represent them as feature vectors. Similar images are retrieved based on metrics such as Euclidean distance or angular cosine. A major advantage is that Resnet18 is available as a pre-trained model so that it can be used for image recognition without being trained from scratch.

Architecturally, Resnet is composed of residual blocks that allow easy propagation of information through the network. By its structure the algorithm solved the problem of information loss in a multi-layer neural network [8]. This phenomenon is called the gradient vanishing problem. In the process of back propagation, the gradient (a measure of the sensitivity of the neural network to changes in its weights) shrinks, causing the initial layers not to train properly which would lead to poor network performance [9].

As a result, the residual blocks contain skip connections that allow information to reach the previous layers.

Images downloaded from external sites were passed through the ResNet-18 convolutional neural network to extract features. Feature vectors were indexed thanks to the Spotify Annoy library which allows approximate nearest neighbour search for points in n-dimensional space.

In order to extract feature vectors, the model was modified by removing the last layer which was responsible for classification. It was provided with 512 features, identified as inputs by the last convolutional layer, and it used them to identify one of the 1000 possible classes. Thus, the 512 features will be provided to the index which based on the angular cosine metric will be able to identify the N nearest neighbours

Index creation is an independent process that can be done only once and shared between different resources. Any image uploaded within the FurnishMe software solution is transformed into a feature vector using the ResNet-18 neural network. The loaded index provides a list of nearest neighbours using the angular cosine method.

4. Experiment & Results

For the development of the furniture object detection model, a series of steps were taken, such as data collection, data processing, training object detection model and prediction. The original YOLOv5 deep learning model can be used free of charge in a pre-trained version on the MS COCO dataset. The MS COCO (Microsoft Common Objects in Context) dataset is widely used for detecting objects belonging to 80 different classes. The FurnishMe application aims to detect furniture objects belonging to 8 classes: bed, sofa, chair, table, curtain, mirror, lamp, cabinetry. The dataset mentioned above includes only half of the classes needed by the computer system (bed, sofa, table, chair). Therefore, the YOLO neural network requires an explicit training process on a custom dataset.

Training a model from scratch is a difficult and time-consuming task. Therefore, I chose transfer learning method which reuses a previously learned model for solving a similar new task. In transfer learning, a machine exploits knowledge acquired from a previous task to improve the generalisation of a new one. In computer vision, neural networks are built in layers that progressively identify features such as lines, colours, textures or shapes, finally reaching the identification of specific features for classification. In the transfer learning technique, layers that identify common features are reused and those used for classification are retrained [10].

Data preparation is another complex and time-consuming process, because YOLO neural network requires the training dataset to be defined in a specific format. For each image in the dataset, a text file with the same name is created. Each text file contains the annotations on the image's item bounding boxes. The structure of a line in the text file consists of:

- object class - an integer representing the object class code. The class index must start at 0 and increase by 1 for each unique class in the dataset;
- X_{centre} , Y_{centre} the coordinates of the centre of the bounding box normalised to the width and height of the image;
- W, H - length and height of the bounding box normalised to image length and height

The image annotation process can be done manually using specialised software such as LabelImg. In this paper, images of interest were taken from a public source of labelled data, Open Images.

A training dataset (80% of the obtained images) and a test dataset (20% of the retrieved images) were created. The training set is used to teach the model how to detect furniture objects. The test dataset is used to assess the model's ability to generalise and predict on unknown data that was not part of the model training process.

The Google Collaboratory service was chosen to train the model thanks to its GPU

which speeds up the training time of a deep learning model. The model was trained for 100 epochs with the following results:

```

Validating runs/train/Model/weights/best.pt...
Fusing layers...
Model summary: 157 layers, 7631701 parameters, 0 gradients
Class      Images  Instances  P      R      mAP50  mAP50-95  mAP50-95:100% 42/42 [00:21:00:00, 1.991tr/s]
all       2659    8468      0.573  0.545  0.514   0.212
Bed       2659    488       0.752  0.789  0.738   0.471
Couch    2659    491       0.586  0.615  0.589   0.353
Chair    2659    2930     0.495  0.585  0.458   0.244
Table    2659    1512     0.477  0.474  0.484   0.214
Cabinetry 2659    1181     0.383  0.355  0.263   0.137
Lamp     2659    789      0.582  0.489  0.439   0.212
Curtain  2659    753      0.534  0.565  0.5     0.29
Mirror   2659    324      0.77   0.688  0.718   0.554
Results saved to runs/train/Model
    
```

Figure 5. YOLO training results

Figure 5 shows the performance metrics. The Precision (P) is represented by the proportion of correct positive predictions relative to all positive predictions (correct or incorrect) made by the model [11]. This metric is useful if we want to reduce the number of false positives, i.e. cases where the model predicts that an object is in the image but in fact it does not exist.

The Recall (R) metric is the ratio of the number of correct positive classifications to the sum of the number of correct positive predictions and the number of incorrect negative predictions [11]. This metric is useful when we want to minimise the number of false negatives, i.e. cases where the model does not identify an object in the image even though it exists [12]. Thus, the P metric shows the accuracy in classifying objects as positive, and the R metric measures the ability of the model to detect positive objects.

According to the Figure 5, the model recognizes an object properly in 58% of cases and it identifies correctly the label of all furniture objects in 55% of cases. For the IoU (Intersection over Union) metric, it is necessary to choose the acceptable threshold for which the model prediction is positive and correct. Usually, the accepted threshold is 50%. By progressively calculating the Precision and Recall metrics for different thresholds, we obtain the Precision-Recall curve visible in Figure 6. We thus observe that as the IoU threshold is lower, the Recall metric which show the ability to detect positive objects increases, but the accuracy decreases.

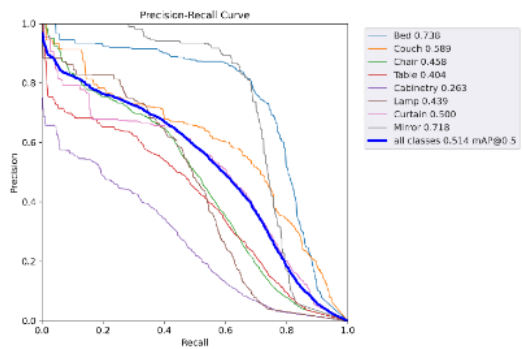


Figure 6. Precision-Recall Curve

Mean Average Precision (mAP) is another important metric in evaluating an object detection model. This metric is a summary of the Precision-Recall curve into a single value that represents the average of all accuracies across classes [11].

Overall, the model exhibits average performance, with moderate metric values for most classes. We note the increased accuracy for labels “bed” and “mirror”. The mAP50 score indicates the model's ability to detect objects consistently with an overlap of more than 0.5 and with an accuracy of about 52%. However, these scores could be improved by optimising the model or increasing the training dataset.

To identify possible errors in the classification process, we analyse the confusion matrix shown in Figure 7.

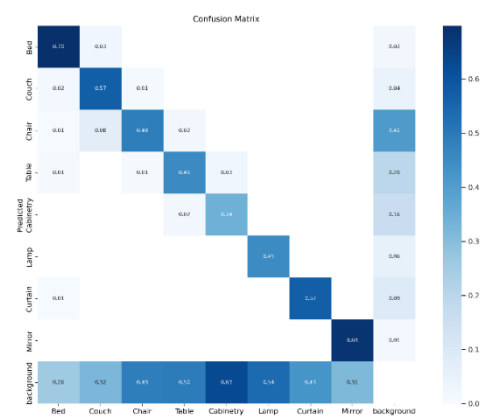


Figure 7. Confusion matrix

The classification accuracy for each class is shown on the diagonal of the confusion matrix. The values range from the minimum accuracy value, recorded for the cabinetry class (0.34), to the maximum accuracy

value, recorded for the bed class (0.7). The confusion matrix reveals a low level of misclassification between furniture items. For example, in some cases the sofa is classified as a chair. The tendency to categorise objects from the classes of interest to the background class has been highlighted as a drawback in the model. This may signal a lack of balanced training data or inadequate architecture.

The purpose of the FurnishMe software solution is to demonstrate the feasibility of a concept by developing a prototype software solution. Due to the limited open-source resources and the model's average performance with moderate metric values, we chose to use it

Since the ResNet-18 neural network was only used for feature vector extraction (and not for classification), the pre-trained version was chosen. In this way, we obtained meaningful image representations without engaging in extensive training. Such a neural network stands out for its adaptability thanks to training on a diverse dataset.

The Mask R-CNN model was used for instance segmentation process. Due to the lack of a dataset including the object masks of the FurnishMe application's eight classes, we used the pre-trained model.

Figure 8 shows the integration of the modules described above in a Flask web application.

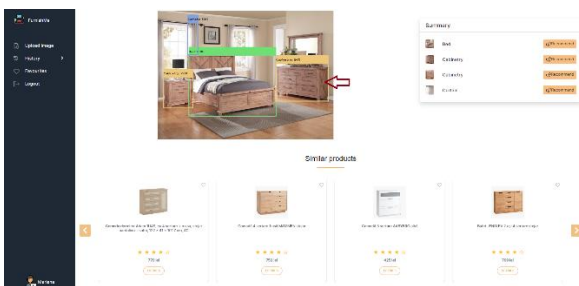


Figure 8. FurnishMe application

5. Conclusions

The main aim of this paper was to develop an innovative information technology system capable of providing users a personalised and easy home design experience. The solution is based on computer vision technology and it allows the

users to explore the furniture of some of the most popular retailers in Romania: Ikea, Jysk and Dedeman. FurnishMe detects furniture items in user-uploaded pictures and recommends visually similar products from the three big shops mentioned above. The application adds value to the field under review by providing an efficient and convenient way for consumers to choose the right furniture. The software solution does not restrict itself to the products of a single retailer, but rather presents the products of Romania's most popular and accessible retailers. This aspect allows the users to analyse and evaluate products based on their specific needs, such as pricing, personal preferences, product availability in a particular store location and so on.

At the moment, the solution is only a prototype that can be extended and improved through additional research. The primary goal of this study was to demonstrate the feasibility of an idea. One aspect that would add value to the software solution is optimising the detection and recommendation algorithm's performance using various model regularisation techniques. Furthermore, user loyalty for the FurnishMe software solution could be strengthened by collecting user preferences and generating personalised recommendations.

Appendix

- CNN – Convolutional Neural Network
- YOLO – You Only Look Once (Object Detection Model)
- NMS – Non-Maximum Suppression
- IoU – Intersection Over Union
- ResNet – Residual Neural Network
- SC-NLMs - Structure-Content Neural Language Models
- LSTM – Long short-term memory
- SPM – Spatial Pyramid Matching
- BoVW – Bag of Visual Words
- LightGBM – Light Gradient Boosting Machine

References

- [1] S. Ozturkcan, "Service innovation: Using augmented reality in the IKEA Place app," *Journal of Information Technology Teaching Cases*, vol. 11, no. 1, pp. 8-13, 2021.
- [2] I. Tautkute, T. Trzciński, A. P. Skorupa, Ł. Brocki, and K. Marasek, "Deepstyle: Multimodal search engine for fashion and interior design," *IEEE Access*, vol. 7, pp. 84613-84628, 2019.
- [3] A. Yaguchi, K. Ono, E. Makihara, N. Ikushima, and T. Nakayama, "Multi-Scale Feature Fusion for Interior Style Detection," *Applied Sciences*, vol. 12, no. 19, pp. 9761, 2022.
- [4] T. Diwan, G. Anirudh, and J. V. Tembhurne, "Object detection using YOLO: Challenges, architectural successors, datasets and applications," *Multimedia Tools and Applications*, vol. 82, no. 6, pp. 9243-9275, 2023.
- [5] H. Liu, F. Sun, J. Gu, and L. Deng, "Sf-yolov5: A lightweight small object detection algorithm based on improved feature fusion mode," *Sensors*, vol. 22, no. 15, pp. 5817, 2022.
- [6] X. Wang and J. Song, "ICIoU: Improved loss based on complete intersection over union for bounding box regression," *IEEE Access*, vol. 9, pp. 105686-105695, 2021.
- [7] T. N. Pham, V. H. Nguyen, and J. H. Huh, "Integration of improved YOLOv5 for face mask detector and auto-labeling to generate dataset for fighting against COVID-19," *The Journal of Supercomputing*, pp. 1-27, 2023.
- [8] E. Limonova, D. Alfonso, D. Nikolaev, and V. V. Arlazarov, "ResNet-like architecture with low hardware requirements," in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 6204-6211, Jan. 2021.
- [9] D. Zou, Y. Cao, D. Zhou, and Q. Gu, "Gradient descent optimizes over-parameterized deep ReLU networks," *Machine Learning*, vol. 109, pp. 467-492, 2020.
- [10] G. Pinto, Z. Wang, A. Roy, T. Hong, and A. Capozzoli, "Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives," *Advances in Applied Energy*, pp. 100084, 2022.
- [11] R. Padilla, W. L. Passos, T. L. Dias, S. L. Netto, and E. A. Da Silva, "A comparative analysis of object detection metrics with a companion open-source toolkit," *Electronics*, vol. 10, no. 3, pp. 279, 2021.
- [12] H. Zhu, H. Wei, B. Li, X. Yuan, and N. Kehtarnavaz, "A review of video object detection: Datasets, metrics and methods," *Applied Sciences*, vol. 10, no. 21, pp. 7834, 2020.



Mariana SERBAN-PREDA – student at Bucharest University of Economic Studies, attending E-Business Master, Bucharest, Romania; obtained a Bachelor's Degree in Economic Informatics in 2021; passionate about Artificial Intelligence, Big Data and Machine Learning;

Data Science Driven Analysis of the Impact of COVID-19

Șerban Gabriel TIMOFTE¹, Ioana Ruxandra STROE², Daniel TAȘCU³,
Raluca Cristina TERTEȘ⁴, Radu Ioan MOGOȘ

^{1,2,3,4,5} The Bucharest University of Economic Studies, Romania

¹EBR Net

timoftesrbn@gmail.com, stroeioana22@stud.ase.ro, tascudaniel22@stud.ase.ro,
tertescristina22@stud.ase.ro, mogos.radu@gmail.com

In the context of a pandemic that emerged with lightning speed, data science has become a cornerstone for governments decision-making processes. By analyzing numerous centralized databases, researchers have been able to identify trends, the spread of the virus, and run artificial intelligence (AI) simulations to anticipate crucial points of the COVID-19 pandemic. Data warehouses created during this period offer real-time monitoring of the global effects of the virus. The health databases are already common in national systems, but their usefulness rises above storing medical histories. The cross-disciplinary nature of the COVID-19 pandemic accentuates the need for collaboration between doctors, medical specialists and data analysts, data engineers, and Artificial Intelligence engineers. This article provides a comprehensive overview of how databases and data warehouses can offer different scenarios for citizens and health specialists alike.

Keywords: Data Science, Data Warehouses, COVID-19, Data Analytics

1 Introduction

Data Science and the health informatics industry are rapidly evolving in the previous years. Health informatics could be described "as an evolving scientific discipline" [1], serving the numerous interest of the specialists - from the research of rare or dangerous diseases (cancer, Parkinson, Crohn) to the prevention of global medical hazards.

Data Science Driven Analysis represents more than running different algorithms on databases. The most important part of a Data Science project is to use proper data sets with real information. Cleaning data represents the process of identifying and removing errors "in order to improve data quality" [2]. However, the process is time-consuming and resource-consuming due to the appearance of "Big Data". This phenomenon led to the appearance of real-time data analysis. In the health system, Big Data is represented by the digital medical records of the patient.

Artificial Intelligence technology was used during the pandemic years, by official entities. For example, World Health Organization (WHO) created a project

called "WHO Coronavirus Dashboard" which is empowered by artificial intelligence software to provide real-time insights around the globe [3]. European Center for Disease Prevention also created during the pandemic a series of AI models to predict the infectious disease. Their model is based on assumptions and it analyses different databases provided by different entities (authorities, hospitals and laboratories). The ECDC's models are focused on how people could interact throughout the day. Also, this is the main reason why are necessary more databases provided by different entities [4]

Access to real information to train the AI models or to create custom dashboards about the spreading of the COVID-19 virus is a problem. Over the internet there are many open-source databases with wrong data and the government does not offer databases in different formats (csv, sql or others). The access is restricted to manual database input. These pandemic problems highlighted the importance of international collaboration to create relational databases with real information.

2. Methods

In our research, we set out to identify how Data Science could impact directly the management of the COVID-19 pandemic and how accessible is this kind of technology. We measured the impact by analyzing the most well-known scenarios about the most affected sections of the population using R and datasets with records around the world and by implementing AI predictions using the SARIMAX model. Pursuing the same concept, we created a Tableau dashboard using real-time synchronization with the database.

Our project is created using the methodologies belonging to multiple sciences: database development, data analytics & science, statistics and machine learning. All of these are “under the hat” of quantitative measures in the digital age. In the following sections, we used open-source databases implemented with PostgreSQL (PSQL), taking care of the data cleaning and normalization process.

The universality of AI technology is demonstrated by world industries. Nowadays, car manufacturers, marketing companies, health equipment, financial technology (FinTech) and many others. The key factor is the knowledge transfer from human to machine. Having a foundation built by human and the capacity of storing large amounts of information, the AI could support people in their activities as an assistant researcher. The machine is able to “memories” significantly more knowledge than a human being, but the human brain is still the cornerstone of society.

An interesting example of implementing AI in a “less popular segment” is presented by doctors Diana Hintea, James Brusey and Elena Gaura in an article regarding the implementation of different AI models to estimate the cabin occupant equivalent temperature [5].

3. Z Tests

For statistical hypothesis analysis, we used the Z test, managing data for all the countries. The Z test result is calculated by dividing the difference between two sample means by the standard error of the difference. This type of hypothesis testing is a standard way of decision-making or a way to analyze the impact of the virus [7]. We consider the null hypothesis (H0) and the HA which rejects the null hypothesis. Also, our risk for this experiment is noted with L.

The first statistical test we implemented examined the hypothesis that “Older people died after the infection with SARS-CoV-2”. All the data are stored in a numeric vector.

```
dead = subset(data, death_dummy == 1)
alive = subset(data, death_dummy == 0)
z.test(alive$age, dead$age, alternative="two.sided", conf.level = 0.99)
mean(alive$age, na.rm = TRUE)
```

Fig. 1. Subsets of Test Number 1

In our test, the H0 is represented by “older people died after SARS-CoV-2 infection”. In this case, the alternative hypothesis is “there is no connection between age and mortality after SARS-CoV-2 infection”. We used the “two-sided” alternative because the scope of the test is to identify if the scenario is true in the provided context.

```
z.test(cured$age, deaths$age,
       alternative="two.sided", conf.level =
       0.99)
```

Fig. 2. Z Test Number 1 in R

The resulting interval is [-25.52122; -15.50661]. Even though the distance between the ends of the interval is far apart, they represent a shred of evidence to confirm the hypothesis. The p-value helps the researcher to determine the statistical significance of the test. In our case, the value of it is 2.2e-16, significantly above 0.05 - the reference value (the alpha from the z.test function body).


```
z.test(men$deaths_dummy, women$deaths_d
       ummy, alternative="two.sided",
       conf.level = 0.99)
```

Fig. 3. Z Test Number 2 in R

For the second test, the resulting interval is $[0,8;8,8]$. It shows the risk of the men dying after the infection. The practical interpretation is that men have a 0,8% to 8% higher risk to die after infection. In this case, the p-value is significantly higher than in the previous test. The value is 0.002, but is still lower than the reference value. However, the mortality of the men is 8,5%, while the mortality of women is 3,7%. These two tests confirm both null hypotheses and we can conclude that the risk is a variable that depends very much on gender and age. Also, in our analysis there is not recorded the impact of the medical records of the patients from the previous years.

4. Summary of the database using Jupyter Notebook

Our objective in Jupyter Notebook was to discover trends of the virus spreads and to view its geographical expansion.

There are two default manners of analyzing trends evolution according to A. Morgan, A. Amed, D. George and M.Hallett:

- "Overall positive change (higher highs and higher lows - expressed as value increase) +1"
- "Overall negative change (lower highs and lower lows - expressed as value decrease) -1" [8]



Fig. 4. Spread map in T0

Considering January 2020 as being the start moment of the pandemic period in the world, we have in the Fig. 4 the map of COVID-19 spread in the debut of the

pandemic. The main variable of the targeted trend is the number of confirmed cases.

The second and the third key moments visible on the world map created in Jupyter Notebook are 03/03 and 03/29. From these two pictures, we can observe that the spread direction is from East (E) to West (W) and in less than a month, all the Europe's countries jumped over the superior limit of our analytics.



Fig. 5. Spread map in T1 - 03/03



Fig. 6. Spread map in T2 - 03/29

In the Fig. 6, the trend is confirmed spreading the virus all around the globe.

In the generated Jupyter Notebook with the evolution of COVID-19 in the world, the user could easily observe that a crowded area with a high number of people in one square kilometer is affected first.

These observations are obvious looking at a video or trying to remember the first part of 2020. However, these videos and analyses would not have been possible without Big Data or Health Informatics. The speed of development is a high advantage in crisis situations when the authority's reaction should be almost immediate. The data analyses role is to manage all the data flow in real-time to help in the management of the decisions.

A clear illustration of the importance of these technologies in a crisis situation is represented in Fig. 7. To be able to visualize the map from Fig. 4, Fig. 5 and Fig. 6 could prevent the authorities and the population that the situation represented in Fig. 7 is imminent.

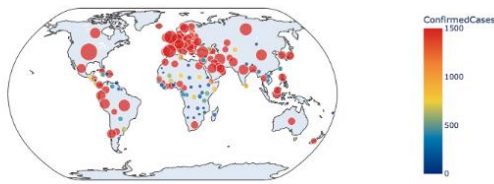


Fig. 7. Spread map in T3 - 05/05

5. Time series

”Time series forecasting models are used to predict the futuristic outcomes based on historical information” [9].

We applied the SARIMAX model to the official datasets of governments about COVID-19 confirmed cases.

Seasonal AutoRegressive Integrated Moving Average with exogenous variables (SARIMAX) is a time-series model based on the popular AutoRegressive Integrated Moving Average (ARIMA) model. The difference is that in the SARIMAX model, the seasonality and exogenous variables are analyzed, too. ”It is interesting to think that all exogenous factors are still technically indirectly modeled in the historical model forecast” [10]. For our analysis, there are four main categories of external data with a significant impact on the prediction accuracy:

- *Geographic & Demographic distribution*
- *The economic power of the state*
- *Urban mobility*
- *Vaccination rate* (this one is not possible to be applied in our research, because at the moment of data recording, there was no vaccine)

In order to apply the SARIMAX model to the COVID-19 dataset, an important step before is a stationary check. The mathematical formula for this verification is that standard deviation is not dependent on time. This is the equivalent of the idea that a shift in the series does not create any change in the shape of its distribution. [11] The Seasonality component of SARIMAX models refers to events that are not

consistent during the time and appears just in shorter periods. Viktor Mehandzhiyski has a practical example of this phenomenon: the ”Jingle Bells” song [12]. In the SARIMAX model, there is used a combination of autoregression with seasonality. The actual output is the dependence of the past values of the variables and the moving average (dependence of the variable on forecast error) [13].

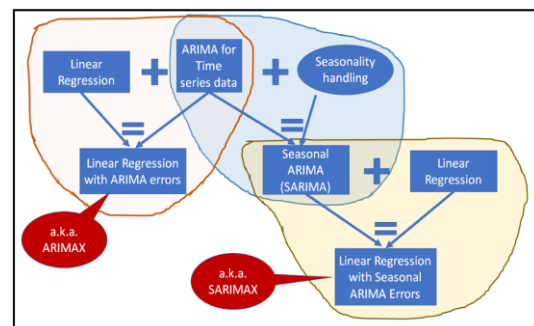


Fig. 8. SARIMAX model components - Time series Analysis, Regression and Forecasting with Python -

<https://timeseriesreasoning.com/contents/regression-with-arima-errors-model/>

In order to apply the SARIMAX model to the COVID-19 dataset, an important step before is a stationary check. The mathematical formula for this verification is that standard deviation is not dependent on time. This is the equivalent of the idea that a shift in the series does not create any change in the shape of its distribution. [11] The Seasonality component of SARIMAX models refers to events that are not consistent during the time and appears just in shorter periods. Viktor Mehandzhiyski has a practical example of this phenomenon: the ”Jingle Bells” song [12]. In the SARIMAX model, there is used a combination of autoregression with seasonality. The actual output is the dependence of the past values of the variables and the moving average (dependence of the variable on forecast error) [13].

The most important principle in which the SARIMAX model works is that training data prepare the algorithm for the prediction. However, the accuracy of the model is measured using the computation with the real situations.

After we trained the model with data from the first three months of the COVID-19 pandemic in Romania and Spain, we collected the next results (Tabel. 1 for Spain and Tabel. 10 for Romania). The datasets were provided as excels from the official websites of the Ministry of Health in these two countries.

Every row in these tables represents the simulation in a day. There is a prediction on 5 days, storing data from the last three months (90 days). The report is 1/18 (simulated days / analyzed days).

Table 1. Romania's prediction

Spain - Prediction	Spain - Reality	Error
188699	170537	-10,6498 8829
189501	174621	-8,521311 87
190303	179143	-6,22966 0104
191104	185870	-2,81594 6629
191906	191444	-0,24132 3834

Intriguingly, the first error values considered in the module are closely (10,65 for Spain and 12,03 for Romania). Looking for other simulations, we observed that the first value of the error in the module is situated between 10% and 15% (Germany 12,43%, Italy 9,8%, France 14,74%). This fact could indicate a standard error of the

model caused by the lack of a larger dataset. Also, SARIMAX is a model that takes into consideration historical data and this is another factor or this first error.

I considered these two countries in my example in order to show the importance of a proper dataset when we are running AI-driven simulations.

Table 2. Spain's prediction

Ro - Prediction	Ro - Reality	Error
2076	2360	12,03389 831
2097	3183	34,118755 89
2121	3502	39,43460 879
2148	3613	40,54802 104
2230	4010	44,38902 743

Spain offered a higher database and the testing ration of the local authorities is higher. In the

Spain scenario, SARIMAX identified correctly the increasing trend of confirmed Covid-19 cases. Looking at the Romanian's prediction, the trend is also correctly identified (in the context of the global raising of confirmed cases). However, the speed of raising is totally wrong. After, we started with an error of 10,6498%, on the last simulated day, Spain had the module value of error around 0,2413. In Romania, the first error is 12,033%, but the last one is 44,389%. These differences are caused by improper datasets. However, the speed of raising is totally wrong. After, we started with an error of 10,6498%, on the last simulated day, Spain had the module value of error around 0,2413. In Romania, the first

error is 12,033%, but the last one is 44,389%. These differences are caused by improper datasets.

	Coefficients	Standard Error	t Stat	P-value
Intercept	633524,324	222605,326	2,84595312	0,10447256
X Variable 1	-1,1198923	1,0224286	-1,0953257	0,38766929

Fig. 8. Error trend

This shows a significant problem of AI-driven simulations. If the researcher and/or the authorities could not provide real data, the output will contain important mistakes. "If the input data is flawed or inaccurate, then any results will be similarly problematic. The quality of data used for machine learning is critical for both the accuracy and reliability of the algorithmic output" [14].

To support the importance of transparency in publishing realistic data, we forward the example of China to analyze. The error is constantly -98%.

6. Vaccine results using regression

The regression model is the cornerstone of any prediction model. The SARIMAX model has an autoregression component included. Regression represents the statistical method to identify the correlation between two variables (the independent and the dependent variable) [15, 16]. The objective of this part of the research is to identify the connection between vaccinations and confirmed cases in a time frame of three months.

The analysis dataset incorporates all the available data on the official websites created by Romanian Government. X is called the dependent variable and in our case, it represents the number of confirmed cases per month starting from July 2021. Y (independent variable) means the number of first doses administrated by Romanians starting from May 2021. The correlation is powerful and it is a negative one. The regression index (r) is 0,6123. According to the coefficient of Intercept and X Variable 1, one more vaccinated person could lead to a decrease in confirmed cases with 1,11.

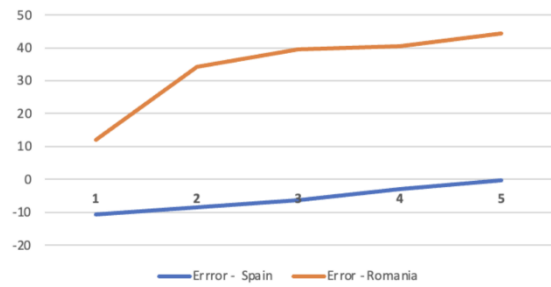


Fig. 9. Regression table (vaccine-confirmed cases)

This result led us to an intermediate conclusion - the vaccine has a small role in stopping the virus spread in Romania, based on the available data.

The resulting regression equation is:

$$Y = 633524,324 - 1,11989 * X$$

The Significant F indicator shows us the relevance of the test. In our case, it is approximately 0,3 and shows us that there are not enough data in the set. We can use our results, taking into consideration that are not accurate enough to reflect totally the truth [17].

The second dataset contains the number of vaccines and the number of deaths caused by coronavirus. The timeframe is similar, comparing the number of vaccines from May and the number of deaths in July. The correlation is weaker this time with $r = 0,469$ (weighted negative connection), but it is still relevant for the macro-analyze. According to the regression coefficients (Intercept and X Variable 1), one more vaccinated person could decrease the number of deaths after two months with 33,46 people.

In this second test, the Significant F has still a high value (0,4). The available data are still not enough in Romania to generalize risk about the phenomenon under investigation.

The resulting regression equation is:

$$Y = 597826,2 - 33,461592 * X$$

Connecting these two results, it is clear that the variables have a connection and the vaccine directly affects both the confirmed cases and the number of deaths. However, this analysis raised yet another question - for one more vaccinated person, the other 33 people do not die because they get infected with a lighter form or not at all?

7. Results

The theories about how age and gender could impact the mortality of coronavirus are true. Both indicators have an important role in treatment.

Data Analytics and Machine Learning are important and efficient tools for risk management. Looking at the provided graphs and simulations, central authorities could prepare the next steps to prevent overcrowding of the health system and reduce mortality. Also, the data analytics field could impact the transparency of the information regarding to impact of coronavirus. Tableau and PowerBI are two important tools for synthesizing big data sets based on geographical locations or the number of cases. These summaries and highlights will support the press and local authorities in their management decisions. The Machine Learning SARIMAX model is also an important way to predict the evolution of the pandemic in a country if a proper database is provided.

The vaccine could reduce the number of deaths in a medium-term (at least 2 months or more) and could decrease the number of confirmed cases.

8. Conclusions

It is clear that Data Science could positively impact the management of crisis situations (in our case - COVID-19 pandemic). By introducing AI models and data analysis techniques, the responsible authorities could increase their efficiency. „AI technologies offer significant opportunities to enhance COVID-19 management and control. Machine learning models and

natural language processing (NLP) techniques are capable of detecting, diagnosing, and predicting COVID-19, which helps public health officials and medical professionals make informed decisions and allocate resources accordingly” [18].

The most significant hindrance is that numerous countries have not published enough data. Most developed countries like France, Spain, the USA or the UK uploaded a significant dataset about confirmed cases and the number of vaccines. In their case, a scientific research is relevant and could support the decisions of governments. We submitted the analysis of Romania, where access to data represented a problem during the pandemic years [19].

Data Science is a pillar of support for Health Informatics and, taking into consideration the best practices, it could positively affect the development of health system around the world!

References

- [1] E. Hovenga, M. Kidd, S. Garde, C. Hullin Lucay Cossion, *Health Informatics*, IOS Press, Amsterdam, Berlin, Tokyo, Washington DC, 2010
- [2] E. Rahm, H. Hai Do, *Data Cleaning: Problems and Current Approaches*, Bulletin of the Technical Committee on Data Engineering, IEE Computer Society, 2008.
- [3] World Health Organization, WHO Coronavirus (Covid-19) Dashboard, 2020
- [4] European Center for Disease Prevention, Monitoring and reporting data and trends, 2020
- [5] Diana Hintea, James Brusey, Elena Gaura, *A Study on Several Machine Learning Methods for Estimating Cabin Occupant Equivalent Temperature*, Coventry University, 2020
- [6] O. J. Watson, *Global impact of the first year of Covid-19 vaccination: a mathematical modelling study*, The Lancet Infection Diseases, 2022

- [7] C. Osborn, *Statistical Applications for Health Information Management, Second Edition*, American Health Information Management Association, 2006
- [8] A. Morgan, A. Amed, D. George, M. Hallet, *Mastering Spark for Data Science*,
- [9] N. Kumar, S. Susan, *Covid-19 Pandemic Prediction using Time Series Forecasting Models*, IEEE Explore, 2020
- [10] B. Artley, *Time Series Forecasting with ARIMA, SARIMA and SARIMAX: A deep-dive on the gold standard of time series forecasting*, TowardsDataScience, 2022
- [11] T. Khachatryan, *Time Series Forecasting with SARIMAX*, Geometrein.medium, 2022
- [12] V. Mehandzhiyski, *What Is a SARIMAX Model*, 365 DataScience, 2023
- [13] Box, G. Jenkins, G. Reinsel, G. C. Ljung, *Time series analysis: forecasting and control*, John Wiley & Sons, 2015
- [14] E. Topol, *The importance of data quality for machine learning in healthcare*, International Journal of Epidemiology, 2020
- [15] D. Montgomery, E. Peck, G. Vining, *Introduction to Linear Regression Analysis*, Wiley, 2012
- [16] P. Patel, *Introduction to Quantitative Methods*, Harvard University, 2009
- [17] D. Paulson, *Handbook of Regression and Modeling*, Chapman & Hall/CRC, 2006
- [18] M. Rahimi, M. Shamsi, A. Rajabi, *Artificial Intelligence applications in managing and controlling Covid-19 pandemic*, Journal of Cellular and Molecular no. 3, 2021
- [19] S. Dascalu, O. Geambașu, C. Raiu, D. Azoicai, E. Popovici, C. Apetrei, *COVID-19 in Romania: What Went Wrong?*, Frontiers, 2021
- [20] R. Klement, H. Walach, *SEIR Models in the light of Critical Realism - A critique of exaggerated claims about the effectiveness of Covid-19 vaccinations*, Futures 148, 2023



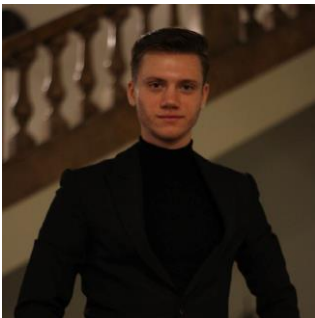
Serban-Gabriel TIMOFTE is student at the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies from 2022, after he studies one year of Computer Science at Coventry University. He worked for international organizations in Romania: Endava, Digital Nation and UK company RWSBC. At present, he works for EBR Net SRL in Bucharest as a software developer and is responsible for the finance department of the start-up. The main focus is to implement AI-related technologies in different business models to anticipate production results and decrease the pollution level of the business.



Ioana Ruxandra STROE graduated from Carol 1 National College in Craiova. Currently, Ruxandra is a first-year student at the Bucharest Academy of Economic Studies, specializing in Economic Informatics. During high school, she was involved in numerous volunteering activities both within her school and outside of it. For example, she helped renovate the school library, one of the oldest in the country, and contributed to the preservation of numerous books and paintings dating back hundreds of years. Additionally, she participated in various volunteering activities organized by JCI, one of the most well-known NGOs dedicated to young people in Craiova.



Cristina Raluca TERTEȘ graduated from Mihai Viteazul National College in Bucharest. Currently, Raluca is a first-year student at the Bucharest University of Economic Studies (ASE), enrolled in the Faculty of Cybernetics, Economic Informatics, and Statistics. Her specialization lies in Economic Informatics, highlighting her interest in applying technology within the economic and business context. Moreover, during high school, she actively participated in a variety of volunteering activities organized by Proedus, the largest volunteer program in Bucharest.



Daniel-Valentin TAȘCU is a first-year student at the Faculty of Cybernetics, Statistics, and Economic Informatics in Bucharest, is driven by a strong desire for knowledge. He excels in his studies and actively volunteers with the student organization, "Business Organization for Students." His dedication to continuous learning and selfless service fuels his ambition to make a positive impact in his field.



Radu-Ioan MOGOȘ is a senior lecturer/associate professor within the Department of Economic IT and Cybernetics at Bucharest University of Economic Studies (ASE). He is teaching disciplines like Programming fundamentals, Programming Techniques and Algorithms, Artificial intelligence Economic Information Systems, Evolutionary Programming and Genetic Algorithms. He is also researcher in the field of applied computer science, being the author or co-author to several books and articles that were published in national and international journals and conference proceedings. He had postdoctoral studies at ASE during 2014-2015 and had defended his PhD thesis on 2011 at ASE. He has a Master Degree in Business Relations and Communication (ASE, 2007) and a master degree in English Language Education and Research Communication for Business and Economics - EDURES (ASE, 2014). He is also member of the Romania Project Management Association, being during the time member in several major projects.

Internet of Things - Overview

Andreea MIHAI¹, Ștefania Codruța MĂNĂILĂ², Antonio Sebastian DUMITRAȘCU³
^{1,2,3}The Bucharest University of Economic Studies, Romania
mihaiandreea17@stud.ase.ro, manailastefania17@stud.ase.ro,
dumitrascuantonio18@stud.ase.ro

Abstract: *The Internet of Things (IOT) is a paradigm that has changed the traditional way of living into one in step with technology. IoT has brought great changes in several fields such as agriculture, energy, healthcare, transportation and infrastructure. A lot has been done to improve IoT technology, but there are still study challenges (technical, political) that need to be solved to reach its full potential. The main purpose of this article is to provide an overview of what IoT means, its evolution and applicability in day-to-day life. This article discusses several aspects such as IoT architecture, IoT challenges, IoT applicability areas, importance of Big Data analytics in IoT and its evolution in the last few years. Furthermore, programming languages that can be used to create IoT-type software and small examples are presented, comparisons between them. This article will help the readers to better understand IoT, real-life applicability, evolution and overview of how to develop an IoT program using Arduino or Raspberry PI.*

Keywords: *Internet of Things (IoT), IoT architecture, IoT challenges, IoT applicability areas, importance of Big Data analytics in IoT, IoT evolution, Arduino, Raspberry PI*

1 Introduction

The term "Internet of Things" (IoT) was coined by Kevin Ashton during a presentation to Procter & Gamble in 1999. The Internet of Things is a subject still under development with technical, social and economic importance. Various consumable products, durable goods, automobiles, industrial machinery, and other common objects are being combined with Internet connectivity and advanced data analysis mechanisms with the promise of changing the way people live and work. With the large-scale deployment of IoT devices, a major change can be seen in people's daily routines. IoT is used everywhere, globally, often without people even realizing it. For consumers, new IoT products such as home process automation devices and electricity management devices have introduced the concept of "Smart home", which has become an emblem of safety and sustainability. Other devices

such as health or fitness monitoring tools are transforming the way medical services can be delivered. The Internet of Things transforms physical objects into a distributed information ecosystem between both fixed and portable devices with the goal of improving the quality of human life. IoT technology also promises to be useful to the elderly or disabled, improving their level of independence at a reasonable cost. Systems such as smart cars, smart traffic orchestrators, sensors embedded in roads and bridges, make it possible to implement the concept of "smart cities", whose main objectives are to reduce traffic and energy consumption. IoT technology also has the potential to transform agriculture, industry, energy production and distribution.

A considerable number of companies and research organizations have speculated on the potential impact of IoT on the economy over the next decade. Huawei expects 100

billion IoT connections by the year 2025. Manyika et al. estimates the potential economic impact of IoT between 3.9 and 11 quadrillion dollars annually starting in 2025, impact caused by: low device prices, advances in cloud computing, and high Internet speed. IoT is expected to contribute between 4% and 11% to global GDP by the year 2025. [1]

However, implementing the Internet of Things raises significant obstacles that could stand in the way of fulfilling its beneficial purpose. Among the risks involved in implementing IoT systems are the hacking of devices, concerns about the invasion of personal space, but skepticism associated with the idea of continuous surveillance. Also, in addition to potential security issues, other obstacles to the large-scale implementation of IoT are technical, political and legal in nature. Thus, this discussion of ‘benefit vs. risk’ along with the flow of information spread through social media make IoT a complex subject to understand.[2]

2. IoT architecture

2.1 IoT architecture based on 4 levels

The layered architecture of the Internet of Things is responsible for the collection, management, storage and processing of data resources. It incorporates 4 main levels, each playing a critical role in the successful implementation of an IoT-type system. There are other secondary levels, lower levels of complexity that help improve system performance.

The IoT architecture represents the configuration of a functional Internet of Things system, in which each component is responsible for its own set of tasks. The IoT architecture offers the systems that implement it flexibility in the process of defining their own characteristics. IoT systems are used in a wide range of

applications in almost all market segments.[3]

The basic architecture of the Internet of Things is composed of four main levels:

- Perception – The Perception level is the level where data is collected by the sensors of the devices connected to the system;
- Connectivity – The Connectivity level is responsible for data transfer and facilitates communication between devices and the system network;
- Processing – The Processing level is the level where the data resources generated by the previous levels are preprocessed and stored;
- Application – The Application level represents the level where data is analyzed and used to facilitate different business requirements such as control systems, advanced analysis, etc.

2.1.1 Perception

The main objective of the Perception layer is to transform digital / analog signals. This layer forms the foundation of the IoT infrastructure, collecting data, but at the same time being able to perform actions on the collected data. Devices connected to the Internet are the bridge between the digital plane and the physical plane.

Many types of devices, using different operating systems, are used within this tier. These devices can be grouped into three categories:

- Sensors – Sensors can be represented in different forms such as meters, pressure gauges or probes. The sensors take various data related to the environment such as temperature, pressure or wind speed and transform the information into digital signals. These signals are then transmitted to cloud platforms for processing;
- Actuators – Actuators play the opposite role of sensors. They transform the

digital signals transmitted by an IoT system into concrete actions that can be used to control the behavior of connected devices;

- Devices – Sensors and actuators are often connected to different devices. In some cases, the sensor or actuator may be the main component of the device.
- The Internet Of Things architecture presents no limitation on the type of device used within the Perception layer, nor on their proximity or geographic location. A device can be a sensor of miniature dimensions or a component of an industrial machine. They can be located at a fixed point or distributed among millions of access points in different locations.[4]

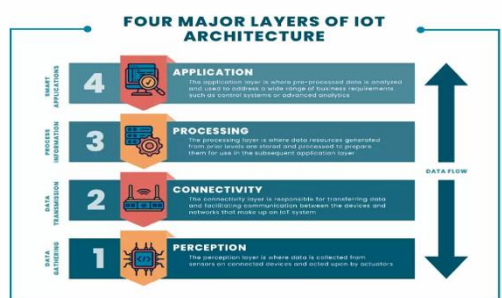


Fig. 1. IoT architecture based on four levels

2.1.2 Connectivity

The connectivity level is responsible for the communication between devices, networks and cloud systems necessary for the architecture of an IoT-type solution. The connection between the Perception layer and the Processing layer is made using the TCP/IP or UDP/IP stack. It can also be implemented using gateways that transform signals to various protocols. With the help of the gateways, it is possible to encrypt / decrypt the data transmitted within the system.

Within this level, multiple communication technologies can be used. The choice of technology used is often dependent on the

system design. Other important factors that influence the implementation of the level of connectivity are the type of devices used, the distance traveled by the signals, but also the potential obstacles that can prevent data traffic.

Among the technologies used within this level are the following:

- Ethernet – Ethernet uses cable communication to provide secure and high-speed connectivity over short distances;
- WiFi – WiFi provides wireless connectivity for short distances;
- Low-power Wide-area Network (LPWAN) – LPWAN is a technology created for IoT devices that provides low power consumption, high battery capacity and wireless connectivity for long distances;
- Cellular networks – Using cellular networks, the IoT system benefits from stable connectivity with global coverage. It involves high cost and high consumption of electricity.
- The main communication protocols used to facilitate data transport between devices and cloud platforms are:
 - Data Distribution Service (DDS) – Ensures the communication of IoT components in real time;
 - Message Queue Telemetry Transport (MQTT) – It is used in collecting data from devices with low energy consumption;
 - Constrained Application Protocol (CoAP) – It is used for devices with resource constraints;
 - Advanced Message Queuing Protocol (AMQP) – Supports the exchange of information between servers.

The optimal combination of technology and communication protocol is essential for creating an IoT architecture.

2.1.3 Processing

The Processing level is the level where the data generated within the Perception level and transmitted through the Connectivity level are collected, stored and processed. The data goes through two stages of processing within this level.

The data accumulation stage is the mechanism by which the generated data is transformed into an interpretable form for the Application layer. The purpose of this stage is to efficiently sort and store the accumulated data, making it accessible to the next level of the system. Data can be stored in different ways depending on its relevance within the system. Multiple technologies are used for IoT data persistence.

The data abstraction stage is the final stage of data preprocessing where it is beautified and abstracted for further use in applications. Data from both IoT and non-IoT sources is unified and brought to one single format. The information is then aggregated and stored centrally to facilitate quick accessibility.

The Processing layer also manages the reformatting of the data returned by the Application layer so that it can be interpreted by the devices.

2.1.4 Application

The Application layer of an IoT infrastructure is where data is analyzed to solve business requirements and meet objectives. Software tools transform data generated by the design layer into meaningful information for users or the system. APIs are used to integrate IoT software with a middleware. Certain IoT platforms offer native data visualization or analysis features.[5]

Users interact directly with the Application layer and can monitor IoT devices and track data through various tools such as

dashboards or mobile applications. This level is realized in different forms in the various areas of domains:

- Mobile phone GPSs are an example of a user-oriented IoT application;
- IoT systems can provide business intelligence solutions regarding consumer activity and market trends to facilitate enterprise decision-making;
- Robots and automated machines are also monitored and controlled through the application layer. Systems can also use machine learning techniques to increase their performance and accuracy as they are used;
- Intelligent road management systems rely on the Application layer of the IoT architecture to provide accurate information.

2.2 IoT architecture based on 7 levels

In addition to the four main layers, additional layers are used in the seven-layer IoT architecture to improve the performance of the IoT infrastructure:

- Edge or Fog computing;
- Business;
- Security;

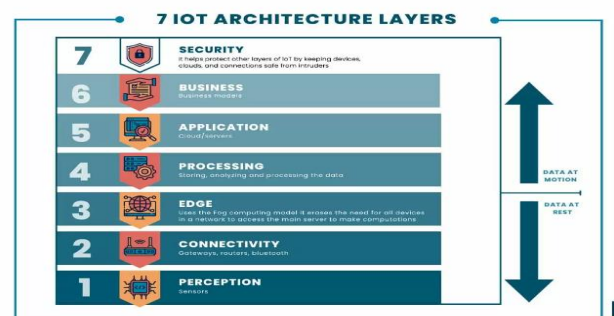


Fig. 2. IoT Architecture with 7 layers

2.2.1 Edge or Fog computing

Edge or Fog computing is designed to reduce latency and take advantage of the availability of 5G mobile networks. The objective of this level is to process and store the information as close as possible to its source, without involving its transmission

to the cloud platform. These activities are performed at the edge of the network to save time and resources needed to communicate with the cloud environment.[6]

Edge computing enables real-time processing of a considerable volume of data. In some cases, the Edge layer plays the role of both the Processing layer and the Application layer, allowing data to be transmitted directly to devices. It can also be used to analyze and evaluate data if it requires further processing or needs to be sent to another point.

2.2.2 Business

The Business layer is where solutions and decisions are drawn from the Application layer. It can contain multiple instances of the Application layer increasing its complexity and arguing the need to separate it into a distinct entity. It is the most important level for enterprises opting for an IoT infrastructure.

2.2.3 Security

E2E security is essential in any IoT infrastructure. Often, confidential data is also transmitted within a system, making security a critical component of the architecture. Compromised devices can put users' health and safety at risk. The Security layer spans all other layers of the architecture and is critical to the success of implementing an IoT solution.

Security can be sub-classified into three main categories:

1. Device Security – Security must be ensured from the very beginning, within the Perception layer. Thus, the devices must be equipped with both hardware and firmware fuses. Specific device security features include:

- Ensuring authentication by means of cryptographic keys;
- Physical protection of devices by strengthening their outer layer;
- Preventing unauthorized code from running on a connected device;

- Installing firmware to address possible vulnerabilities.

2. Connection security – Data must be secured by means of encryption at the time of its transport within the Connectivity layer. End-to-end scripting eliminates the risk of data being intercepted and used by unauthorized users.

3. Cloud Security – Cloud security is achieved by encrypting stored data to reduce the risk of disclosure of confidential information involved in a possible security breach. Access to IoT applications must be limited by implementing a set of measures aimed at authorizing and authenticating users. Devices also need to be authorized when connecting them to the IoT system to prevent unauthorized devices from connecting.[7]

Security must be the main feature in all IoT deployments. Their complexity and flexibility can often involve difficulties in securing the infrastructure. All possible security issues must be resolved before the system is put into production.

3. IoT challenges

The Internet of Things (IoT) is often hailed for its revolutionary potential in terms of application, which is why IoT devices are increasingly used in various economic sectors for varied purposes. However, there are still uncertainties about how the IoT will evolve, as it has to face some technical challenges as well as some political challenges.

3.1 Technical challenges

There are certain technical limitations that influence the development and applicability of IoT applications, such as the absence of Internet addresses within the most widely used protocol, the limited availability of high-speed wireless communications, and the lack of consensus on technical details.

3.1.1 Internet addresses

There is a significant limitation in IoT development, which lies in the technical restrictions of the Internet Protocol (IP) version most commonly used today. Internet Protocol consists of a set of standards used by computers to transmit and receive information over the Internet, including the unique address each device must be connected to in order to communicate. The predominant version of the IP protocol is IPv4, which allows for around four billion addresses, and this capacity is almost exhausted, with only a few new addresses available in various regions of the world.[8] Forecasts indicate rapid growth in Internet traffic over the next five years, largely driven by IoT technologies compared to other devices that require Internet connectivity.[9] In 2020, out of a total of approximately 50 billion devices connected to the Internet, over 25 billion were IoT devices.[10] Despite some workarounds, such as IP address sharing, it is unlikely that IPv4 will be able to meet this growing demand.

Conversely, the advent of version 6 of the IP protocol can support a significant increase in the number of available IP addresses. There is a strong likelihood that IPv6 will be widely deployed to handle the expected increase in the number of devices connected to the Internet. IPv6 has been available since 1999, but was only officially launched in 2012. By 2015, in most countries, less than 10% of IP addresses were using this version. Its spread is highest in certain European countries and in the United States of America, where it has seen an increase of up to 20% in recent years. Globally, the adoption of version 6 has doubled annually since 2011, reaching approximately 7% of all addresses in mid-2015. However, continued growth in adoption does not guarantee that IPv6 will be able to meet the evolution of IoT device usage. This depends on several factors, such as updating legacy systems and applications that cannot support IPv6 addresses,

addressing security issues associated with the transition, and the availability of development resources.[11]

3.1.2 High speed internet

It is obvious that IoT devices fundamentally require an internet connection to function effectively. This requirement can become a challenge if the Internet connection is not of high quality and telecommunication services are not advanced enough.

Although Internet access is generally available in urban and suburban environments, the situation is different in rural environments, where private service providers have difficulty installing the necessary infrastructure, as it is not considered profitable and government programs are limited.

3.1.3 Wireless communications

According to experts in the field, it is considered essential to solve the issues related to access to the electromagnetic spectrum to ensure efficient operation and interoperability of IoT devices. Access to spectrum, whether licensed or not, is crucial for wireless communication of devices and objects without the need for a physical connection. Although IoT devices are used in various industries and purposes, there are opinions that the current allocation of the electromagnetic spectrum is not adequate to meet the specific needs of IoT solutions in these fields and industries.

3.1.4 Technical standards

Currently, there is no global consensus on technical standards for the Internet of Things (IoT), especially regarding communication-related aspects, and no universally accepted definition in organizations responsible for developing standards or common documents for IoT.[12] However, the need for a set of common standards and rules is recognized to ensure the interconnectivity and

scalability of IoT devices and systems. However, there are concerns about adopting common development standards, due to the diversity of IoT devices and the specific needs of different industries and fields.[13]

3.1.5 Other technical issues

In addition to the existence of a common implementation standard, we also face other technical challenges that can limit the development of IoT applications. For example, the security and safe updating of smart objects can be an issue. It is recommended that IoT devices provide remote update capability to ensure proper security and functionality. However, there is a risk that these upgrade capabilities may have unwanted effects, such as increasing the power requirements of IoT devices or requiring the implementation of additional functionality to ensure security during the upgrade process and to avoid the risk of an attack by hackers.

Energy efficiency is a significant technical challenge when it comes to IoT devices. The functionality of these devices relies on the consumption of energy to sense, process and communicate information. Especially if the devices are located off the grid and use batteries, replacing or recharging them can become a problem, even when optimizing energy use. This problem is commonly encountered in applications involving a large number of devices or hard-to-reach locations. As a result, alternative solutions are being developed to obtain electricity, such as the use of solar or other renewable energy sources.

3.2 Political challenges

3.2.1 Privacy

Privacy in IoT is essential to prevent abuse and unauthorized disclosure of data. Significant efforts have been made to identify privacy issues and find appropriate solutions to protect both user data and the devices themselves. Bandyopadhyay and

Sen (2011) [14] contributed with some relevant solutions in this area:

- a) People usually do not agree with the idea of their data being accessible to the general public. Therefore, it is necessary to have a control over the personal information of users to ensure the confidentiality of their data.
- b) There is a strong desire that people not be tracked without their consent. However, to protect location privacy, it is necessary to have adequate control over the user's physical location and movements.
- c) In order to guarantee the right to confidentiality, it is necessary to implement standardized policies at the legislative level in this regard.
- d) To ensure effective privacy management, it is essential to develop regulated standards, methodologies and tools.

According to previous studies (Chan and Perring 2003) [15], data associated with users and devices in the IoT must be stored in authorized servers and accessible only by authorized individuals or entities. In the context of IoT systems, various entities communicate and interact, each with their own privacy policies. For this reason, conflicts and inconsistencies between these policies inevitably arise. It is thus necessary to develop new solutions to check, notify and resolve these consistency issues. Given that privacy policies are a real limitation to the effective interaction of IoT systems, there is a strong motivation for researchers to create a common and descriptive language that defines a standard set of privacy policies.

3.2.2 Security

Privacy and security are interrelated issues in the IoT field. Security is a significant challenge for physical IoT components such as wireless sensor networks (WSNs) and radio frequency identification (RFID) devices. This is due to limited resources, computational constraints, limited storage

spaces, low battery power, and other factors. As networks and sensor devices are used in sensitive applications, they are at risk of attacks in insecure environments. Any attack on one node in an IoT system can compromise the entire network of sensor nodes.

Software and hardware improvements are aimed at preventing such events in certain situations. However, to effectively manage security issues, advanced countermeasures such as vulnerable node detection techniques, encryption algorithms, encryption key management mechanisms, and secure routing protocols are required. These additional measures contribute to increasing the level of security and protection of IoT systems.

Thus, it can be seen that privacy and security are essential aspects in the field of IoT and require appropriate approaches and solutions to ensure data protection and systems functionality.

In a restricted environment where there are valuable objects or equipment, it is important to have applications that provide immediate notifications to users if they are moved or taken without permission. These applications may use various communication methods, such as SMS text messages, emails, or voice recordings, to inform the user of the unauthorized event.

Even in the presence of the risk of security attacks, IoT applications and services must be able to perform their functions and retrieve data in real time. This involves implementing appropriate security measures such as user authentication, data encryption, and activity monitoring to detect suspicious behavior.

In the event of unexpected security attacks, devices or nodes may need to be reprogrammed to fix the system. This process may involve updating software, applying security patches, or changing the configuration of devices to prevent exploitation of vulnerabilities.[16]

3.2.2.1 Data authentication

Data authentication in IoT communications is essential to avoid situations where data sent by a sender does not reach the specified receiver or to prevent unauthorized modification or interception. One of the techniques used to authenticate messages is the Hash Message Authentication Code (HMAC), according to Krawczyk (1997). [17]

HMAC is a cryptographic technique that uses a shared secret key between the parties involved in the communication. This key is used in conjunction with a hash function, such as MD5 (Message Digest Algorithm 5) or SHA (Secure Hash Algorithm), to calculate a message authentication code. This authentication code is attached to the message and can be verified by the receiver of the message using the same secret key and the corresponding hash function.

By using HMAC, messages can be securely authenticated, guaranteeing that they come from the stated source and have not been altered during transmission. The technique provides an additional level of security and integrity within IoT communication, ensuring that data is authentic and has not been compromised during transfer.

It is important to note that as technology advances, other methods of authenticating data within the IoT have been developed, such as digital certificates and advanced cryptographic protocols. These techniques can be tailored to the specific needs of the IoT system and can provide high levels of security and authentication.

3.2.2.3 Data integration

In the context of data integration in IoT systems, it is crucial to ensure the integrity of messages and verify the veracity of their sources. RFID (Radio Frequency Identification) tags, commonly used in IoT, present certain challenges in this regard, as they are often unsupervised and can be vulnerable to attack.

According to Jules (2006) [18], RFID tags may be at risk of data alteration when transferred over the network. This can be a major concern from a security perspective, as the data stored and generated by RFID tags can be manipulated by attackers, which can compromise their integrity and authenticity.

To protect the data stored in RFID tags against attacks and memory corruption, various memory protection techniques have been proposed. For example, EPCglobal class-1 and generation-2 tags use passwords to protect memory against read and overwrite operations. The memory of such a tag is divided into layers, and each layer is protected independently by using a password.

However, using passwords presents certain challenges. Often, RFID tags only support short passwords, which can compromise security. In addition, managing communication between different IoT systems becomes complex when they belong to different organizations and must rely on interoperable authentication and security mechanisms.

Consequently, ensuring the security and integrity of data in IoT systems, including the use of RFID tags, requires the implementation of robust solutions and protocols. These may involve advanced methods of encryption, authentication and access management, as well as common standards and agreements between the various systems and organizations involved in the IoT.

3.2.3 Trust and governance

Within the IoT domain, there is no general definition for the concept of trust. However, according to Blaze (1996) [19], trust can be defined as a set of policies and credentials used to manage access to resources.

According to Roman's (2011) study [20], in the IoT environment, trust mechanisms must meet certain essential requirements:

- Lowering the level of insecurity of objects when they interact with each other;
- Coordination of objects in associating with reliable partners for the fulfillment of the same purpose;
- Ensuring reliable, dynamic and cooperative environments in which the objects can carry out their activity;
- Understanding the effects that IoT systems can cause on users' sensations when they interact with them. Users must have control over their own services and must have the necessary tools to accurately describe their interactions with the IoT world;

Daubert (2015) [21] classifies trust into four categories:

- Device Trust: Supports the need to interact only with trusted devices. To achieve this purpose, reliable software products and schemes must be used;
- Trust in the process: implies the need to interact only with relevant and correct data. In this case, accurate data collection, sustainable data analysis and data fusion are required;
- Connection trust: involves the need to exchange information only with the right provider service. Here, data integration, authentication and privacy processes are required.
- Trust in the system: implies the need to interact in a general system of trust. Here the workflows must be provided in a transparent manner, the processes and technologies underlying the system and also described within the specific contexts.

4. Areas of applicability of IoT

IoT has the potential to significantly impact various sectors of the economy and society. However, the extent and nature of its evolution remain uncertain. The development of IoT is expected to bring positive benefits in terms of integration, efficiency and productivity across multiple sectors and global economies. Agriculture,

energy, healthcare, manufacturing services, and the transportation sector are among those most likely to benefit from IoT advancements. This development could have a favorable impact on economic growth, infrastructures and cities, as well as on ordinary consumers. However, there are technical and political challenges that can slow IoT development and trust, including security and privacy issues, as mentioned earlier. [22]

Various economic analyzes indicate that the IoT will contribute significantly to economic growth over the next decade, but estimates vary considerably as to the magnitude of this impact. The current IoT market has been valued at approximately two trillion dollars, and in the next five to ten years it is forecast to reach a value between four and 11 trillion dollars. The wide variation reflects the difficulty in making economic forecasts, as there are various uncertainties and no common trajectory accepted by researchers regarding the exact definition of the IoT and how it will develop in the future. Next, it's described how the influence of IoT can be felt in different economic sectors. [23]

4.1 Agriculture

Agriculture is adopting IoT technology to achieve accurate results, optimize production and efficiency, while reducing costs and the impact of climate events. IoT enables detailed analysis of real-time data collected in agricultural processes, including climate conditions, air and water quality, soil condition, water supplies, pest population, crop development, and other factors such as costs, equipment availability, and labor management.[24] For example, soil sensors are used to precisely monitor soil moisture and chemical balance.[25] These sensors can be integrated with technologies installed in different areas of crops to activate precise irrigation and fertilization. Drones and satellites are used to provide clear images of land quality, giving farmers information on

crop progress, nutrient deficiencies and the location of weeds. In animal husbandry, technologies such as radio frequency identification (RFID) chips and electronic identification tags (EID) are used to monitor animal movements, feeding and breeding patterns and generate detailed individual reports for each animal. [26]

4.2 Energy

In the energy industry, IoT technology is having a significant impact on energy production and distribution, facilitating the control of oil pipelines and wellheads, for example. When IoT components are integrated into the various parts of the power grid, an infrastructure called the "smart grid" is formed. This IoT integration enables more precise control of electricity flow and can improve the efficiency of network operations. It also facilitates the integration of microgenerators into the grid. Smart grids allow ordinary consumers to have greater control over energy consumption in homes and offices. The installation of smart meters in temperature or lighting control systems, as well as in other systems in a building, leads to the emergence of "smart buildings" that integrate the operation of these systems. Smart buildings rely on the use of sensors and data to regulate temperature, lighting and overall energy consumption, resulting in reduced energy waste and lower costs for consumers. Information collected from buildings located in a narrow area can later be integrated to achieve additional efficiency at the level of a neighborhood or a wider area of a city. [27]

4.3 The health system

IoT solutions are making a significant impact in the healthcare industry, focusing on the monitoring and treatment of conditions by providing remote medical services. These applications use medical technologies and the Internet to provide remote healthcare and education. IoT

devices, such as wearables, implantables, or ingestibles, provide information about vital signs, chronic conditions, and other indicators of health and well-being. Thus, the healthcare sector benefits from a wide spectrum of IoT devices adapted to individual needs. [28]

4.4 Manufacturing

In the industrial sector, the implementation of IoT solutions is revolutionizing manufacturing processes and supply chain logistics. Optimizing operations is one of the main benefits these applications bring, turning factories into more efficient units. By connecting components in the factory, production can be optimized, and by connecting components in the inventory and distribution process, the supply chain can be optimized. In addition, IoT applications used in this field facilitate maintenance and enable the prediction of potential failures. Sensors monitor machinery and factory infrastructure, and the data collected is used by maintenance teams to decide when to replace faulty parts, thus avoiding unwanted and costly events. [29]

4.5 Transport

In the field of transport, interconnectivity is becoming more and more present. Most motor vehicles today are equipped with GPS applications and entertainment systems, as well as driver assistance systems that use sensors for maneuvers such as parking or emergency braking. With the advancement of IoT, it is estimated that in the next 5-20 years, fully autonomous vehicles will be commercialized, capable of operating without human intervention, relying only on the interconnectivity of the vehicle's systems. [30]

In addition, IoT technologies enable the interconnection of vehicles of various categories, including cars, buses, trains, airplanes, and drones, to form an Intelligent Transportation System (ITS). This system

facilitates communication between vehicles and IoT infrastructure components, resulting in the prevention of accidents, optimization of traffic flow, saving of energy resources and reduction of associated costs. [31]

4.6 Infrastructure and smart cities

IoT infrastructure can be implemented in various sectors of public utilities, including sewage, water transport and treatment, public transport and waste management, helping to create an integrated and efficient infrastructure, especially in the urban environment. For example, traffic management authorities can use cameras with built-in sensors to monitor and regulate traffic flow, thereby reducing congestion. IoT components integrated into public lighting or other elements of street infrastructure can provide advanced lighting control functionality, environmental monitoring and easy parking assistance. In addition, smart waste containers can communicate with public cleaning teams when they are full, thus optimizing collection routes. [32]

In the context of smart cities, the integration of infrastructure components with services has led to the development of a complex concept. There is still no standard definition or rules that clearly characterize such a city, as the characteristics can vary considerably. In general, the smart city concept refers to the use of IoT technologies to improve energy consumption, transportation processes, governance and other public services in order to meet specific needs and promote sustainability and improve the quality of life. Technologies IoT communicates with include: [33]

- Virtual social networks, such as Facebook or Twitter, which facilitate interaction between city residents and local authorities, as well as the exchange of information and opinions.
- Mobile devices, such as phones and smart watches, which allow easy access to information and services, such as

interactive maps, public transport schedules or mobile payment services.

- Data analysis, which involves the processing and use of large volumes of information (the phenomenon of big data) and the use of public and widely accessible databases (open data) to obtain relevant information and make more informed decisions regarding urban planning, resource management or energy efficiency.
- Cloud computing, which involves the distribution of computing and storage services to a remote infrastructure, thus offering the possibility to access and use resources in a flexible and scalable way.

5. Importance of Big Data analytics in IoT

Big Data plays an important role in the field of IoT (Internet of Things) as it enables organizations to collect, analyze and interpret large amounts of data generated by IoT devices. IoT has also revolutionized the way we live and work, and this aspect refers to the network of physical devices, vehicles, buildings and other elements embedded with sensors, software and network connectivity that enable them to collect and exchange data. The high volume and variety of data generated by these devices presents significant challenges for analysis and decision-making. This is where big data analysis comes in.

Big data analytics refers to the process of analyzing large and complex data sets to uncover hidden patterns, correlations, and insights. Some of the key reasons why big data is important in IoT are: [34]

1. Data collection and management: IoT devices generate large amounts of data, which can be difficult to manage and analyze without the use of big data technologies. Using big data tools and platforms, organizations can effectively collect, store and manage data from IoT devices.

In general, there are three main stages of the data collection and management process in IoT:

a) Data collection - This involves capturing and collecting data from IoT devices such as temperature, humidity or pressure sensors. This data is often collected through wireless communication technology such as Wi-Fi, Bluetooth or cellular networks.

b) Data processing - After data is collected, it must be processed before it can be used. This process involves processing raw data into an usable format, such as converting the data to digital format, and then applying algorithms and analysis models to generate valuable insights.

c) Data management - The processed data is then stored in a centralized system such as a cloud server. This data must be managed in a secure and reliable way to avoid its loss or corruption. This involves the use of security technologies and tools to protect data against cyber-attacks and other threats.



Fig. 3. Big Data characteristics

In addition, there are several aspects of data collection and management in the IoT that need to be considered, including:

Large volume of data - With billions of connected IoT devices, the volume of data generated is enormous. This can lead to difficulties in data management and processing.

Data Quality - Data collected from IoT devices can be inaccurate or incomplete, which can affect the value of the information obtained through data analysis. It is important to ensure good data quality

by using high quality sensors and by validating the data before processing.

Data Integrity - Data collected in IoT must be managed in a secure and reliable way to avoid its loss or corruption. This involves the use of security technologies and tools to protect data against cyber-attacks and other threats.

In general, data collection and management in IoT are critical processes to be able to get the maximum benefits from this field.

2. **Real-time analysis:** Big data technologies enable real-time analysis of data, which is essential for IoT applications that require real-time decision making. For example, in a smart city application, real-time analytics can help city officials respond quickly to traffic congestion or other issues.
3. **Predictive analytics:** IoT devices can collect data about their own performance and usage, which can be used to predict when maintenance is needed. Big data analysis of historical data collected from IoT devices can help organizations develop predictive analytics models to analyze this data to identify patterns and predict/anticipate and prevent potential equipment failures before they occur, reducing downtime and costs. For example, this type of predictive analytics can be applied in manufacturing.
4. **Improved customer experience:** By analyzing big data from IoT devices, organizations can gain insights into customer behavior and preferences, which can be used to provide insights and recommendations based on large and complex data sets. For example, retailers can use data from IoT devices to better understand customer behavior and preferences, allowing them to develop their products, personalize their services and improve the customer experience, leading to a stronger relationship between companies and customers.
5. **Enhanced Security:** IoT devices can be vulnerable to cyber-attacks, and big data analytics can help organizations detect and respond to these threats in real time.

By analyzing data from multiple sources, organizations can identify patterns and anomalies that may indicate a security breach.

Data security in IoT is essential to protect confidential information and prevent data loss or corruption. By implementing appropriate security measures, the protection of sensitive data can be ensured and the maximum benefits of big data analytics in IoT can be obtained.

An IoT system comprises a large number of devices and sensors that communicate with each other. With the expansion of IoT, the number of sensors and devices is growing rapidly. These devices communicate with each other and transfer large volumes of data over the Internet. This data is very large and is transmitted every second and hence it is called "big data". The continuous expansion of IoT-based networks gives rise to complex issues such as data management and collection, storage and processing and analysis. The large amount of IoT data for smart buildings is very useful to deal with several issues of smart buildings, such as managing oxygen levels, measuring smoke/hazardous gases and brightness. Such a framework is able to collect data from sensors installed in buildings and perform data analysis for decision making. In addition, industrial production can be improved using an IoT-based cyber-physical system that is equipped with information analysis and knowledge acquisition techniques. Traffic congestion is an important issue in smart cities. Real-time traffic information can be collected through IoT devices and sensors installed in traffic signals, and this information can be analyzed in an IoT-based traffic management system. In healthcare analytics, IoT sensors used with patients generate a wealth of information about patients' health every second. This large amount of information must be integrated into a single database and must be processed in real time to make quick decisions with high accuracy, and big data technology is the best solution for this job. IoT along with

big data analytics can also help transform the traditional approaches used in manufacturing industries into a modern one. Sensing devices generate information that can be analyzed using big data approaches and can help with various decision-making tasks. In addition, the use of cloud computing and analytics can benefit the development and conservation of low-cost energy and customer satisfaction. IoT devices generate a huge amount of streaming data that needs to be stored efficiently, additional analytics for real-time decision making. Deep learning is very efficient to deal with such a large amount of information and can provide results with high accuracy. Therefore, IoT, Big Data Analytics and Deep learning together are very important to develop a high-tech society.

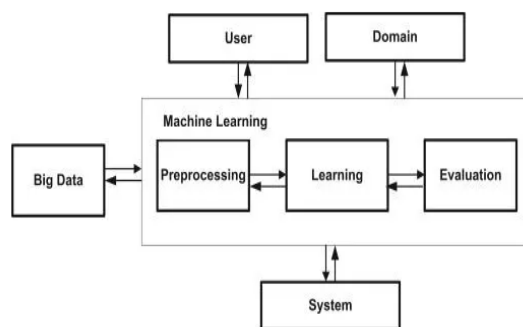


Fig. 4. IoT, Big Data and Machine learning interaction

In conclusion, big data analytics is critical to IoT as it enables organizations to extract meaningful insights from the large amounts of data generated by IoT devices. It enables organizations to monitor, analyze and respond to real-time data, predict potential problems, improve decision-making and enhance security.

6. The evolution of IoT in the last few years

In the last few years, the Internet of Things (IoT) has experienced a significant evolution, being considered one of the most important trends in the field of technology.

One of the key factors contributing to the evolution of IoT has been the development of connectivity technology such as 5G and Wi-Fi 6, which have increased the speed and capacity of data transmission. This has allowed IoT devices to communicate more efficiently and collect and transmit more data in real time.

In addition, the evolution of sensor technology has led to the development of smaller and more efficient IoT devices that can be used in various fields such as health, agriculture or manufacturing. For example, sensors can be used to collect data on the health status of patients, monitor crops and improve the efficiency of production processes.

Another important trend in the evolution of IoT is the increased use of data analytics, especially predictive analytics and artificial intelligence. These technologies enable IoT devices to collect and analyze data in real time, providing valuable insights and automated actions such as alerting users or adjusting device settings.

In conclusion, the IoT has experienced significant evolution in recent years due to the development of connectivity and sensor technology, as well as the increased use of data analytics and artificial intelligence. This evolution has enabled IoT devices to become more efficient and useful in various fields, thereby improving the quality of people's lives and the efficiency of industrial processes.

With the evolution of IoT in recent years, this has even started to be taught to children. Teaching IoT to children is an important initiative that helps them understand how technology works and how it can be applied in different fields such as industry, health, agriculture or household. Learning about IoT can give children a wider perspective on the world of technology and help develop their skills in an interactive and creative way.

There are different approaches in teaching IoT to children, such as using games and interactive experiments to make children understand the basic concepts of IoT and

apply them in practical activities. For example, children can build a simple IoT device, such as a temperature sensor, and learn how to collect and analyze the resulting data.

In addition, teaching IoT to children can help develop digital skills such as programming, design and analytical thinking. These skills are important in today's digital age and are needed to face future challenges in career and personal life. Teaching IoT to children can also help develop teamwork and collaboration skills, as IoT projects often require collaboration between different people with different expertise.

In conclusion, teaching IoT to children can have multiple benefits, such as developing digital skills, team and collaboration skills, and their ability to understand and apply technology in different fields.

There are a variety of IDEs (Integrated Development Environments) available for developing IoT applications, which are designed to facilitate the process of programming and testing IoT devices and applications. Below are some examples: [35]

1. Arduino IDE - is an open source platform, used for the development and programming of Arduino microcontrollers. It is a popular and easy-to-use solution for novice developers.

2. Eclipse IoT - is a suite of development tools, including Eclipse Kura, Eclipse Paho and Eclipse SmartHome. Eclipse Kura is an IoT gateway platform and Eclipse Paho is an MQTT client library. Eclipse SmartHome is an open source platform for developing smart home automation solutions.

3. Visual Studio Code - is an open source code editor developed by Microsoft that supports the development of IoT applications. It can be used to develop IoT applications for devices like Raspberry Pi and Arduino.

4. PlatformIO - is an open source platform that supports the development of IoT applications for a wide range of devices,

including Arduino, ESP8266, ESP32 and Raspberry Pi.

5. Particle IDE - is a cloud-based IoT application development platform that supports the development and testing of applications for Particle devices.

These are just a few examples of IDEs for IoT application development. In general, IDEs must support IoT device programming, application development, testing, and debugging. It should also be easy to use and support a wide range of devices and platforms.

Arduino and Raspberry Pi are popular platforms for building electronic projects and creating interactive devices. However, they have some key differences.

Arduino is a microcontroller board that is designed to build simple electronic projects. It has a small form factor and is easy to use, even for beginners. The Arduino board contains a microcontroller chip, which is programmed using the Arduino software. Arduino is great for projects that require precise timing or need to interact with physical sensors or actuators.

Raspberry Pi, on the other hand, is a single-board computer that runs a full operating system (such as Linux). It is more powerful than an Arduino and can handle more complex tasks such as running a web server or media center. The Raspberry Pi also has built-in Wi-Fi and Bluetooth connectivity, making it easy to connect to other devices and the internet.

In short, if one wants to build a simple electronic project that interacts with physical sensors or actuators, Arduino is a good choice. If one wants to build a more complex project that requires computing power, connectivity and the ability to run a full operating system, then the Raspberry Pi may be a better choice.

Raspberry Pi	Arduino
Microcomputer	Microcontroller
Needs an operating system	Does not need an operating system
Complicated	Simple
Video out, Camera, Ethernet ports, Wifi, Bluetooth, USB, I2C, SPI, UART etc. on board	USB only for power and serial in/out, I2C, SPI, UART
Best for general computer	Best for small tasks that constantly repeat
Capable of performing a huge range of tasks	Optimised for sensing and controlling the world around it
Best for more advanced makers	Best for beginners
Programmed in many languages, including C/C++, Python, Ruby	Programmed in C/C++
Relatively high power consumption	Relatively low power consumption


Raspberry Pi Full Stack 

Fig. 5. Differences between Raspberry Pi and Arduino

```
int led=7;
// Funcția de setare (inițializare)
void setup() {
// inițializare pin digital 7 ca o ieșire.
pinMode(led, OUTPUT);
}
// funcția de buclă infinita
void loop() {
digitalWrite(led, HIGH); // pornește (scrie digital)
LED- (HIGH este nivelul de tensiune)
delay(1000); // așteapta o secundă
digitalWrite(led, LOW); // pornește (scrie digital)
LED- (LOW este nivelul de tensiune)
delay(1000); // așteaptă o secundă
}
```

Fig. 6. Script written in Arduino IDE using C programming language

In the academies where programming courses are taught to children, because they do not have the level of thinking/analysis very well developed and the necessary practice, the initiation is started in building such projects using Arduino as a board and Arduino IDE for writing the code.

Example script for lighting an LED written in Arduino IDE for Arduino and Python for Raspberry Pi:

Through the script above, the LED turns on for one second, then stays off for one second. The operation is repeated ad infinitum. [36]

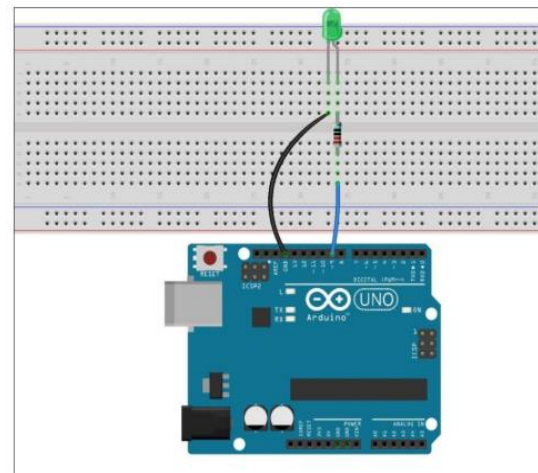


Fig. 7. Hardware schema for lighting a led

```
python
import RPi.GPIO as GPIO
import time

GPIO.setmode(GPIO.BOARD)
GPIO.setup(11, GPIO.OUT)

while True:
    GPIO.output(11, GPIO.HIGH)
    time.sleep(1)
    GPIO.output(11, GPIO.LOW)
    time.sleep(1)
```

Fig. 8. Written script in Python for Raspberry Pi

This code starts by setting the GPIO mode to board mode and configuring pin 11 as an output. It then uses an infinite while loop to turn the LED on and off repeatedly, with a 1 second delay between each state. The LED should be connected between pin 11 and a resistance of about 220 ohms and to ground. [37]

It is important not to connect the LED directly to the Raspberry Pi without a resistor, as this can cause the GPIO pin to burn.

In conclusion, Arduino and Raspberry Pi are two different platforms, each with its own advantages and disadvantages. Arduino is better for simple electronics projects that require precise timing and interaction with physical sensors or actuators, and Raspberry Pi is better for more complex projects that require

computing power, connectivity, and the ability to run a full operating system. In general, Arduino is easier to use and more beginner-friendly, while Raspberry Pi offers more flexibility and more processing power for more advanced projects. Depending on the project, the choice between the two depends on the specific requirements and available resources.

7. Conclusions

Recent technological advancements in the field of IoT expertise have attracted the attention of the international community of researchers and software developers. IoT developers are working together with researchers to expand the scalability of this new type of technology and to potentiate its benefits to society to the highest possible level. However, improvements and progress are only possible if the current unaddressed issues involved in the widespread use of IoT systems are resolved. In this scientific article, both the current context of IoT technology and the problems and challenges encountered by developers in the process of building an optimal Internet of Things model were presented. This technology not only brings benefits to various branches of human activity, but is also a way to gather unlimited amounts of data. Therefore, Big Data Analytics is also a critical factor to consider in order to build an optimal IoT model.

References

- [1] "Internet of Things (IoT)" <https://www.scirp.org/journal/paperinformation.aspx?paperid=108574>
- [2] "IoT Architecture" [Online]. Available: <https://www.javatpoint.com/iot-architecture-models>
- [3] "IoT Architecture Guide. Major and additional layers of IoT system" [Online]. Available: <https://www.helpwire.app/blog/iot-architecture/>
- [4] "What is IoT Architecture?" [Online]. Available: <https://www.mongodb.com/cloud-explained/iot-architecture>
- [5] "IoT Architecture – Detailed Explanation" [Online]. Available: <https://www.interviewbit.com/blog/iot-architecture/>
- [6] "IoT Architecture: Complete Explanation with Examples" [Online]. Available: <https://www.celona.io/network-architecture/iot-architecture>
- [7] "IoT Architecture" [Online]. Available: <https://www.javatpoint.com/iot-architecture-models>
- [8] "It's official: North America out of new IPv4 addresses" [Online]. Available: <https://arstechnica.com/information-technology/2015/07/us-exhausts-new-ipv4-addresses-waitlist-begins/>
- [9] "Cisco Annual Internet Report" [Online]. Available: <https://www.cisco.com/c/en/us/solutions/executive-perspectives/annual-internet-report/index.html>
- [10] "The Internet of Things How the Next Evolution of the Internet Is Changing Everything" [Online]. Available: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf
- [11] "The Internet of Things: Frequently Asked Questions" [Online]. Available: <https://sgp.fas.org/crs/misc/R44227.pdf>
- [12] "A guide to the confusing Internet of Things standards world" [Online]. Available: <https://www.networkworld.com/article/2456421/a-guide-to-the-confusing-internet-of-things-standards-world.html>
- [13] "The state of IoT standards: Stand by for the big shakeout" [Online]. Available: <https://techbeacon.com/app-dev-testing/state-iot-standards-stand-big-shakeout>
- [14] Bandyopadhyay, D. and Sen, J. (2011) 'internet of things: applications and challenges in technology and standardization', Wireless Personal Communications, Vol. 58, No. 1, pp. 49-69
- [15] Chan, H. and Perrig, A. (2003) 'Security and privacy in sensor networks', IEEE Computer, Vol. 36, No. 10, pp. 103-105

- [16] "Internet of Things: A Survey of Challenges and Issues" [Online]. Available: - https://www.researchgate.net/publication/322603292_Internet_of_Things_A_Survey_of_Challenges_and_Issues
- [17] Krawczyk, H., Bellare, M. and Canetti, R. (1997) 'HMAC: keyed-hashing for message authentication', IETF RFC
- [18] [Juels, A. (2006) 'RFID security and privacy: a research survey', IEEE Journal on Selected Areas in Communications (J-SAC
- [19] Blaze, M., Feigenbaum, J. and Lacy, J. (1996) 'Decentralized trust management', IEEE Symposium on Security and Privacy, pp. 164-173
- [20] Roman, R., Najera, P. and Lopez, J. (2011) 'Securing the internet of things', IEEE Computer, Vol. 44, No. 9, pp. 51-58
- [21] Daubert, J., Wiesmaier, A. and Kikiras, P. (2015) 'A view on privacy & trust in IoT', The IEEE International Conference on Communication Workshop (ICCW), pp. 2665-2670
- [22] National Security Telecommunications Advisory Committee, "NSTAC Report to the President on the Internet of Things," November 19, 2014
- [23] "The Internet of Things Will Thrive by 2025" [Online]. Available: <https://www.pewresearch.org/internet/2014/05/14/internet-of-things/>
- [24] "Surprise: Agriculture is doing more with IoT Innovation than most other industries" [Online]. Available: <https://venturebeat.com/business/surprise-agriculture-is-doing-more-with-iot-innovation-than-most-other-industries/>
- [25] "Imagery: Which Way Is Right For Me?" [Online]. Available: <https://www.globalagtechinitiative.com/in-field-technologies/imagery/imagery-which-way-is-right-for-me/>
- [26] "Internet of cows: technology could help track disease, but ranchers are resistant" [Online]. Available: <https://www.theverge.com/2013/5/10/4316658/internet-of-cows-technology-offers-ways-to-track-livestock-but>
- [27] "Smart Grid. Department of Energy" [Online]. Available: <https://www.energy.gov/smart-grid>
- [28] Manyika et al., "The Internet of Things: Mapping the Value Beyond the Hype."
- [29] Macaulay, Buckalew, and Chung, "Internet of Things in Logistics.
- [30] [30] "Autonomous Vehicle Technology" [Online]. Available: https://www.rand.org/pubs/research_reports/RR443-2.html
- [31] "USDOT's Intelligent Transportation Systems (ITS) Strategic Plan 2015-2019" [Online]. Available: <https://www.its.dot.gov/strategicplan.pdf>
- [32] "NAED-Big-Data.pdf" - [Online]. Available: <https://cdn2.hubspot.net/hubfs/2859863/Research%20Report%20Downloads/NAED-Big-Data.pdf>
- [33] "Smart cities" meet "anchor institutions": the case of broadband and the public library. [Online]. Available: <https://www.thefree-library.com/%22Smart+cities%22+meet+%22anchor+institutions%22%3A+the+case+of+broadband+and...-a0403061636>
- [34] "Big Data and IoT: Benefits, Challenges, Use Cases" [Online]. Available: <https://anywhere.epam.com/business/big-data-analytics-and-internet-of-things>
- [35] "Best IoT Development Tools" [Online]. Available: <https://www.g2.com/categories/iot-development-tools>
- [36] "Blink" [Online]. Available: <https://docs.arduino.cc/built-in-examples/basics/Blink>
- [37] "HOW TO CONTROL LEDS WITH THE RASPBERRY PI AND PYTHON" [Online]. Available: <https://www.circuitbasics.com/how-to-control-led-using-raspberry-pi-and-python/>



Andreea MIHAI – student at Bucharest University of Economic Studies, attending Data Bases – Support for Business Master, Bucharest, Romania; obtained a Bachelor’s Degree in Economic Informatics in 2021.



Ștefania – Codruta MANAILA – student at Bucharest University of Economic Studies, attending Data Bases – Support for Business Master, Bucharest, Romania; obtained a Bachelor’s Degree in Economic Informatics in 2021.



Antonio – Sebastian DUMITRASCU – student at Bucharest University of Economic Studies, attending Data Bases – Support for Business Master, Bucharest, Romania; obtained a Bachelor’s Degree in Economic Informatics in 2021.

The Impact of Artificial Intelligence and the ChatGPT Tool on Society. A Case Study on Employees Working in Corporations in Romania

Elena COSTIN¹, Elena Daniela DUMITRAȘCU²,
Andreea Georgiana ENE³, Florinela Georgiana PREDA⁴
^{1,2,3,4} Bucharest University of Economic Studies, Romania
costinlena18@stud.ase.ro, dumitrascuelena21@stud.ase.ro,
eneandreea18@stud.ase.ro, predaflorinela18@stud.ase.ro

Abstract: *This paper aims to study the evolution of artificial intelligence, its impact on education and the job market, and the attitude of a sample of the population towards its potential effects. The study also focuses on the newly emerged chatbot, ChatGPT. For a full perspective on the topic, a survey was conducted to capture the perspective of Romanian employees in corporations regarding the impact of artificial intelligence.*

Keywords: Artificial Intelligence, ChatGPT, Education, Job Market, Corporate Environment

1 Introduction: Artificial intelligence, a highly popular and debated field that is constantly evolving and expanding, seems to be becoming a part of more and more people's lives, sometimes helping them without their conscious awareness. Concrete examples of artificial intelligence applications include virtual assistants, autonomous vehicles, recommendation systems, fraud detection, medical diagnosis, and video games. In this context, the recent emergence of the ChatGPT tool has further amplified the popularity of AI.

Being a highly debated topic, we have observed that there are both positive and negative opinions regarding the impact that artificial intelligence will have on society. Two of the most controversial areas where its presence is felt are education and the job market. While the benefits brought by AI may seem evident, there are also concerns about the consequences of its use in a less ethical manner.

The structure of the paper primarily consists of four parts: introductory aspects regarding artificial intelligence and the ChatGPT tool, their impact on education, their impact on the job market, and a case study. The study data was collected in March 2023, using an online survey with 12 questions, to which 326 individuals responded. The questions focused on both artificial intelligence in

general and the specific aspect of the new ChatGPT tool.

2 What is artificial intelligence?

Artificial intelligence is a field within science and technology that develops intelligent machines and computer programs capable of performing a wide range of tasks that typically require human intelligence. It is a system that mimics various functions that a human can perform. AI uses external data, such as big data, to achieve outstanding performance in the assigned tasks [1].

In his work, [17] provides an answer to the question "What is artificial intelligence?" stating that it is the science and engineering of making intelligent machines, particularly intelligent computer programs. It is related to the similar task of using computers to understand human intelligence, but AI is not confined to methods that are biologically observable.

3 Brief History of Artificial Intelligence

The history of Artificial Intelligence originates from fantasies crafted by philosophers and fiction writers. Philosophers introduced the concept of intelligent machines to help us understand the significance of being human. French philosopher René Descartes employed the

metaphor of a "mechanical man" to illustrate the potential of intelligent machines. Similarly, science fiction writers like Jules Verne (19th century) and Isaac Asimov (20th century) also utilized this possibility to bring fantasy into their works about intelligent non-human beings. Many AI researchers have drawn inspiration from the writings of these authors for their own research [1].

During World War II, English mathematician Alan Turing invented a machine for code-breaking called "The Bombe," which is considered the first functional electromechanical computer. He developed this machine for the British government to decrypt the Enigma code used by the German military during the war [1]. In 1950, Turing published a paper titled "Computing Machinery and Intelligence," in which he discusses the conditions for considering a machine as intelligent, later known as the Turing test. He argued that if a machine could successfully claim to be human to an informed observer, then it should undoubtedly be considered intelligent. The Turing test is a one-sided test. A machine that passes the test should certainly be regarded as intelligent, but a machine could still be considered intelligent without possessing sufficient knowledge about humans to imitate a human [17].

One of the early paradigms of artificial intelligence revolved around high-level thinking [7]. It did not prioritize the capacity to recognize concepts, perceive objects, or perform complex motor skills shared by most animals. Instead, it emphasized the potential to engage in multi-step reasoning, comprehend the meaning of natural language, generate new plans to achieve goals, and even evaluate its own reasoning. This human-like general intelligence was termed strong Artificial Intelligence. The main focus for strong Artificial Intelligence has been on symbolic reasoning, which emphasizes that computers are not just numerical calculators but rather general symbol manipulators. As observed by the authors [22] in their

hypothesis on the physical symbol system, intelligent behavior appears to require the ability to interpret and manipulate symbolic structures. Although this approach initially showed promise, many branches of artificial intelligence have moved away from this approach due to its difficulty and lack of progress in the 21st century. It remains uncertain when and if strong AI will become a reality.

4 Branches of Artificial Intelligence

In his work, [17] identifies some of the branches of Artificial Intelligence, namely:

- Logical AI - involves representing the knowledge about the world, specific situation, and goals of a program using mathematical logic statements. The program uses deduction to determine suitable actions for achieving its objectives.
- Search - AI programs extensively analyze numerous possibilities, whether it's moves in a chess game or inferences in theorem proving programs. Ongoing advancements are constantly being made to enhance the efficiency of these processes across different domains.
- Pattern recognition - When a program makes observations, it is commonly programmed to compare its findings with a predefined model. For instance, a visual recognition program may attempt to match a model of eyes and a nose in order to identify a face.
- Inference - Nonmonotonic inference methods were introduced to logic in the 1970s to handle situations where conclusions can be deduced implicitly but retracted if contradictory evidence arises. These methods allow for reasoning with uncertainty and provide flexibility in drawing conclusions from given facts.
- Common sense knowledge and reasoning - Despite being actively researched since the 1950s, AI still lags behind human-level capabilities in this particular aspect. Despite notable advancements in areas like non-monotonic reasoning systems and theories of action,

further innovative ideas are required to bridge the existing gap between AI and human-level intelligence.

- **Learning from experience:** Programs accomplish this task. AI approaches based on connectionism and neural networks excel in this area. Additionally, there is the learning of laws expressed in logic. However, it is important to note that programs can only learn what facts or behaviors can be represented within their formalisms. Unfortunately, learning systems often have limited abilities to represent information, which poses a challenge in their overall learning capabilities.
- **Planning -** Planning programs utilize general knowledge about the world, specifically the effects of actions, along with situational facts and a defined goal to generate a strategy for achieving that goal. In typical scenarios, the strategy consists of a sequence of actions.
- **Epistemology -** This field of study focuses on understanding the types of knowledge necessary for problem-solving in various real-world scenarios.
- **Ontology -** Ontology involves examining the nature of existence and the different types of entities that exist. In the context of AI, programs and statements deal with various objects, and the field explores the classification and fundamental properties of these object types.
- **Heuristics -** A heuristic is a method or technique used within a program to uncover information or generate ideas. In the field of AI, the term is employed in different contexts. Heuristic functions are utilized in certain search approaches to evaluate the distance between a node in a search tree and the desired goal.
- **Genetic programming -** Genetic programming is a methodology that involves mating random Lisp programs and iteratively selecting the fittest programs over millions of generations in order to solve a given task.

5 Applications of Artificial Intelligence

Artificial intelligence, in the form of neural networks and expert systems, finds applications across various human activities. The combination of high precision and reduced computation time makes AI a cutting-edge technology. It is already transforming industries by replacing humans in workshop-level jobs, allowing them to take on supervisory roles instead. Additionally, brokerage firms now rely on AI to analyze data, perform market analyses, and execute stock transactions without human interference. According to [4], some of the applications of artificial intelligence include: gaming industry, heavy industry, weather forecasts, expert systems, and data mining.

6 ChatGPT – a chatbot based on Artificial Intelligence

ChatGPT, launched in November 2022 by the OpenAI research institute, is a language model that utilizes artificial intelligence to engage in conversational interactions. It is capable of answering questions, acknowledging errors, and challenging incorrect assumptions. The model gained rapid popularity, offering a free version during the research preview that is accessible to everyone. ChatGPT is an AI-powered chatbot platform that utilizes OpenAI's powerful GPT-3 language model. It is highly advanced, capable of generating human-like conversations, and offers a high level of customization. The AI behind ChatGPT is reliable, secure, and ensures context-aware and trustworthy interactions. It is considered one of the most advanced and engaging AI chatbot platforms available [24].

To make this study more interesting, we decided to ask ChatGPT about the algorithms and techniques it uses. The response received was satisfactory, providing clear explanations that are accessible to everyone. The answer can be seen in the following paragraph:

"I have been built based on the GPT-3.5 architecture, which is an enhanced version of the GPT-3 architecture developed by OpenAI. This architecture relies on a combination of several artificial intelligence algorithms and techniques, such as:

- Attention mechanism: This is an algorithm used to understand the relationships between words in a text and identify key words that are relevant for understanding the content.
- Deep learning: This is an algorithm that involves training a model with a large dataset to learn patterns and make predictions.
- Supervised learning: This is an algorithm that involves training a model with a set of input and output data to learn how to make accurate predictions for new input data.
- Unsupervised learning: This is an algorithm that involves training a model without a specific set of input and output data. Instead, it seeks to identify patterns and structures within the input data.
- Transfer learning: This is an algorithm that involves using knowledge learned from one dataset to aid in training a model for another dataset.

In general, multiple artificial intelligence algorithms and techniques are used to create large language models like mine, and they are continuously optimized to enhance the performance and accuracy of the model. ” (ChatGPT, 2023)

7 The Impact of Artificial Intelligence in Education

The Impact of Artificial Intelligence in Education Artificial Intelligence (AI) is the advancement of technology to take over human functions such as the ability to make rational choices, learn, plan and be creative. In the 21st century, technological systems are ubiquitous in social life and society, and its integration in the field of education becomes essential. In recent years, there have been many written studies debating the

impact that artificial intelligence will have in the field of education. The role that ChatGPT can have in adapting the learning system to the individual student, the level of efficiency and the assessment process in higher education [8]. For example, ChatGPT can be integrated into building custom quizzes and assessments, or be useful in grading papers. Using interactivity and capturing students' attention, ChatGPT can create game-based assessments, providing a more engaging way to identify the level of knowledge gained throughout the learning process. The software can be used to streamline the process of grading essays or written assignments. These are just a few examples of how software can be used in the evaluation process, according to the study [8]. Also in this article, the authors also mention a number of methods by which the possible plagiarism generated by the help provided by artificial intelligence can be kept under control. From the article, we can understand on the one hand, the systems that can detect the presence of artificial intelligence within any product, and on the other hand, a series of steps are promoted on how to teach young people to use these tools only as ways of quick access to information, not to plagiarize. According to Study.com's January 2023 study, over 89% of students use the ChatGPT tool for homework, 48% for a test or homework, and 52% of students use it to write essays. The study shows us the positive impact ChatGPT has on feedback on the paper he writes. But all this partner can also be analyzed under the aspect of a negative influence by diminishing or even eliminating the originality or his contribution to the composition of the work [5]. According to the study [30] who mentions the research done by Study.com, ChatGPT should not be used excessively comparing it to different activities where the technology has been introduced and which have not completely replaced the activity, such as : the car and the subway did not and should not replace walking in the park, because physical activity helps our health, just as the electric

bicycle should not be used excessively [30]. [5] presents how teachers can combat artificial intelligence, using the AICheatCheck software, which indicates with 99.7% accuracy the authenticity of the work created by students.

AICheatCheck is a tool built from a model trained on approximately 50,000 human-authored papers spanning multiple domains. The data was divided into a training set consisting of 70% of the respondents, a test set with 20% percentages and a validation set of 10% by which human text was compared with that processed by artificial intelligence. The model at that time had certain limitations, but nowadays this software is constantly being worked on to show as much accuracy as possible and to be a help to the teaching staff [5]. Through the study [2] we understand from the beginning the digital age and the need of the society we live in to align with technology. In the first part, it illustrates the help that society receives through the integration of technology in daily activities. In support of this statement, the author gives examples of smart applications on the market that facilitate the development and performance of actions, such as programs that analyze images, automatic translation programs, voice and facial recognition programs; the artificial intelligence that is incorporated in different products: robots, cars, drones, homes, etc. Also in this article, we are presented with the daily evolution of information technology. Thus, says the author, the rate of evolution of technology must also be in the field of education and the economy, so that young people can have access to them. Moreover, the daily and professional life of man cannot be imagined without this technical-scientific process. Technologies have been around for more than 50 years, and the innovation, updates that are made daily are major advancements for various fields. Artificial intelligence can be considered a central element of the digital age. Artificial intelligence "is seen as a basic element of the digital transformation

of society with a strategic value for education and as a priority" claims the European Union [29].

AI can be used in different stages of the learning process:

- Adapting the learning system to each student - the use of artificial intelligence in the education system could help the student by offering him a way of working that matches his learning model learning. An AI system may be able to identify the student's gaps and adapt to his learning need.
- The level of efficiency of the learning process - the delivery of the learning process can be improved by reducing the time spent analyzing the learning data to identify trends, patterns and adaptation of the learning content of each student.
- Easier access to education - the introduction of artificial intelligence within platforms within learning units in areas where resources are limited.
- Evaluation process - integrating AI into the evaluation process helps reduce costs and time. The integration of AI and related technologies must be done with a focus on people and anchored in human rights, ethics, so that this tool becomes a service, a common good and satisfies a general interest. The implementation in the system of education, culture and in the media sector should follow the fundamental norms and values for a private life, for data protection, intellectual property rights and expression. Exploiting artificial intelligence to an exaggerated extent could influence the cornerstones of social values and fundamental rights, thus the algorithms that will be used, the technologies in which ChatGPT will integrate will have to guarantee maximum protection and a general lack of impartiality [11].

According to various publications, the use of artificial intelligence, since ChatGPT has taken off, entails different aspects of how artificial intelligence can be used, such as:

- Equity - in Romania there are still households that do not have access to the Internet, according to the National Institute of Statistics, in 2021, the access threshold to this resource is below 85% of the population, which generates a digital divide between students who have access and those who do not [9].
- Over technology – over using this tool can lead to misinformation as some of the information they find may not exist. On the "Common Sense Media" website, which is a website that provides resources and advice for parents, the possibility of such a situation is indicated and it is suggested to remove its access to campuses [16].
- Ethics - occurs when users act without filtering the information that the ChatGPT tool. For example, when a company uses artificial intelligence in the selection of candidates, the tool can make decisions that cannot be justified and through which some of the candidates can be discriminated.
- Data Processing and Security – sensitive information about teachers and students may be stored. Information about minor students is stored without parental permission, and in this regard, Italy has blocked access to ChatGPT at present and launched a lawsuit for non-compliance with personal data legislation [12].

We reaffirm that the development of applications that have an AI component should be used in the context of deepening information, because otherwise it will be a risk factor with discriminatory practices, committing crimes or other aspects of a criminal nature. Currently, ChatGPT collects a lot of data, which is a risk factor if that data reaches malicious people. At the educational level, another negative aspect that ChatGPT presents, in addition to the actions that students use for the purpose of discrimination, are those that concern the carrying out of disinformation activities or the promotion of erroneous information by

creating sites in this direction or fake publications.

One of the main advantages of implementing artificial intelligence in schools is that it can help improve the teaching and learning process. For example, artificial intelligence can be used to create personalized learning materials tailored to the individual needs of each student. In addition, artificial intelligence can be used to assess the performance of students and provide them with immediate feedback, which can help improve their performance.

8 The impact of artificial intelligence on the labor market

Artificial intelligence has emerged as a significant catalyst for innovation, permeating various service sectors and delivering distinctive customer experiences through its algorithms [15]. As AI technology continues to proliferate, the apprehension surrounding its potential to supplant human workers has grown substantially. This phenomenon necessitates careful examination, as the ramifications of such a scenario would include widespread job displacement, leaving many individuals without income and necessitating their reskilling to pursue alternative occupations that remain unaffected by technology.

Job insecurity stands as the primary stress factor in the workplace [21], which can even lead to burnout, as there is a positive correlation between the fear of job replacement by artificial intelligence (AI) and burnout. People worry not only about their positions vanishing completely but also about a potential decrease in the overall number of jobs due to AI implementation. Employees who have job security are more committed and capable of providing a positive customer experience. However, those who anticipate future unemployment experience anxiety and frustration, leading to decreased performance [14] and a lack of motivation to continue working in the same position or industry. Additionally, job

insecurity can contribute to various medical complications, both physical and mental, such as heart conditions, insomnia, and general unease [28].

Individuals experience fear when new technologies are introduced within companies, as they perceive that employees with limited technical skills will be replaced [28]. Nevertheless, if they have supportive measures in place, such as employee unions and fair employment contracts, their fear diminishes.

To stay relevant in their respective fields, individuals need to acquire the skills to effectively utilize the new technologies employed in the workplace and be adept at adapting to changes as they occur. Employees are expected to demonstrate high levels of innovation and proactively identify opportunities to generate solutions [23].

Even though artificial intelligence and other new technologies are increasingly utilized and have the potential of partially or fully replace employees in certain professions, they cannot completely replace human creativity in problem-solving. Individuals with this capacity are necessary for revolutionizing other industries and addressing challenges related to food and water, the environment, energy, and security. Innovation involves adapting existing solutions [23]. Another essential attribute for employees is strong communication skills, which are limited when relying solely on artificial intelligence. Moreover, the same studies [23] assert that teamwork is still highly valued in the job market.

The adoption of AI should be considered from the standpoint of service consumers as well, as some people may not prefer to be assisted by a robot [19]. In such cases, technological innovation could have an adverse effect, leading to a reduction in customers or the loss of certain demographics, such as the elderly. It is also highlighted the difficulty of replacing employees with robots in settings like nurseries, healthcare, and education, where

human interaction holds significant importance [19].

Based on a survey conducted in Japan with 10,000 randomly selected respondents, it was found that men have a lower level of fear regarding job loss to artificial intelligence compared to women and younger respondents [19]. Furthermore, respondents who have attended vocational schools and those with higher education working in science and engineering do not perceive AI as a threat to their jobs. They recognize the significance of human interaction in their respective fields.

From the consumers' perspective, 58.6% of them would like artificial intelligence to replace general household chores, while 47.1% would desire robots for the care of the elderly and the sick.

Due to their lack of emotions, empathy, and personal needs, robots may be preferred by companies over human workforce. They do not experience fatigue, require wages, can work beyond regular hours, do not get sick, take leave, or have familial responsibilities. Furthermore, they can operate in hazardous environments that could jeopardize the health of human workers. If robots are able to learn and mimic human emotions, it is expected that vocational jobs will also be replaced for future generations. It is projected that around 3.5 million jobs will be replaced by robots by 2025 [13].

According to researchers [27], the fourth industrial revolution characterized by robots poses a particular threat to women's jobs. In the United States, it is predicted that within the next 20 years, 47% of jobs will be at risk of disappearing due to artificial intelligence [27]. Additionally, studies indicate that in the next 25 years, 10% of driving jobs in the same country will be replaced by self-driving cars [27]. In Japan, significant savings of approximately \$21 billion in elderly healthcare insurance are expected through the replacement of caregivers with robots, while in Australia, miners will be substituted with robotic systems [13]. To address the imbalance caused by job replacement, studies [25]

recommend employers to create more job opportunities that require human skills such as creativity, imagination, and passion.

Artificial intelligence has already been integrated into more than 30 hotels in the United States, including major chains like Sheraton and Hilton. Customer inquiries are handled by chatbots through text or audio, and robots are even providing services such as room service and assistance. The Covid pandemic has accelerated the adoption of contactless services in the hotel industry, aiming to minimize human interaction [15]. AI can analyze customer data derived from their vacation planning activities, such as hotel reservations, payment transactions, and satisfaction surveys, to identify their needs [20]. It is expected that in the hotel industry, jobs involving repetitive tasks and lower skill levels, such as room maintenance, will be automated and replaced by robots. This enables businesses to reduce costs and enhance customer satisfaction. However, in areas where human skills and emotional connection are still essential, AI is projected to serve as a complementary tool. Moreover, this complementarity is more prominent in cases where employees possess higher education qualifications [19].

It is important to highlight that in the case of a mistake made by a human, the blame falls on them, whereas if the mistake is made by a robot, the company for which it works is held responsible [3].

Artificial intelligence also presents some benefits. It can create new jobs [18] and can assist employees in heavily burdened sectors such as the healthcare system to offload some of their tasks.

Recently, the OpenAI-developed ChatGPT has gained popularity, raising concerns about its potential impact on jobs due to its impressive accuracy in responding to questions. What sets ChatGPT apart from other robots is its ability to comprehend contextual information in conversations, enabling more precise solutions. It can be utilized by non-technical individuals through natural language communication in

various languages. Despite the real risk of job displacement, some researchers [10] hold an optimistic perspective on the presence of ChatGPT in people's lives.

An illustrative example of the benefits of ChatGPT can be observed in the field of e-commerce, where this tool has the potential to replace call center staff, leading to cost reductions while assisting customers by streamlining the communication process and replacing traditional methods such as email or phone calls. Customers can receive quick responses, and companies can have better control over the way information is conveyed, which may not be fully achievable in human interactions. Furthermore, an important advantage is the ability of ChatGPT to communicate in multiple languages, surpassing the capabilities of a human employee.

In the field of medicine, patients could utilize systems integrated with ChatGPT to receive medical advice on symptoms and treatment, reducing the wait time for a doctor's response. Artificial intelligence has the potential to analyze medical imaging, diagnose illnesses, and provide specialized treatment recommendations to patients. This would allow doctors to concentrate on more critical cases that require human attention, while routine cases could be efficiently addressed by AI. This approach, as suggested by the authors [10], aims to optimize the allocation of medical resources and improve patient care.

In the field of education, teachers could utilize ChatGPT to provide students with personalized learning plans based on their interests and abilities. Additionally, it could facilitate the workload of teachers by expediting the grading process.

In the banking industry, ChatGPT has the potential to answer customer queries regarding the bank's products, offer personalized solutions based on their needs, and identify potential fraud by recognizing patterns in customer behavior. It can automate various banking processes, predict a customer's loan repayment ability, and analyze transactions and market

conditions to aid in financial decision-making.

Another example highlighted by the authors [10] is the application of ChatGPT in the IT field. They argue that this tool can aid programmers in IT companies by identifying code errors, enabling them to dedicate more attention to developing desired functionalities. Furthermore, it can even locate code snippets in the necessary programming language to resolve specific computer issues.

However, it is important to note that ChatGPT has its limitations, as researchers have pointed out. For instance, it struggles with complex mathematical problems [10]. Additionally, the current interface lacks the ability to upload images or audio files, and there is a tendency for ChatGPT to provide overly detailed responses even when the user requests a simpler answer [26].

9 Case study on the attitude of Romanian corporatists towards artificial intelligence

In order to identify the impact Artificial Intelligence has on Romanians, we made a case study that highlights the attitude of corporate employees towards Artificial Intelligence, focusing on the chatbot recently launched by OpenAI.

Research methodology

Research purpose: Identifying the attitude of Romanian corporate employees regarding the use of artificial intelligence and ChatGPT chatbot.

Research objectives

1. Determining the percentage in which Romanian corporatists know the field of artificial intelligence.
2. Determining the percentage in which the respondents consider beneficial the use of artificial intelligence for everyday activities.

However, the researchers acknowledge that ChatGPT also has limitations, some of which are the inability to correctly execute complex math problems [10] the inability of

the current interface to upload images or audio files, and creating responses much too detailed, although the user mentions at the start of the conversation that he wants an answer as simple as possible [26].

3. Determining the percentage in which the respondents believe that the use of artificial intelligence will facilitate their activities in the workplace.
4. Identifying the percentage of respondents who believe that there will be drastic changes in the labor market, determined by the emergence of artificial intelligence.
5. Identifying ChatGPT chatbot notoriety.
6. Determining the percentage of respondents who rate the ChatGPT chatbot as useful.
7. Determining the level at which the studied population perceives the use of the ChatGPT chatbot as having negative effects on the educational process.

Sample

We chose to study the Romanian population who work in a corporation, because they are more likely to use artificial intelligence in the workplace, and most of the activities carried out involve the use of technology.

To calculate the sample size, the formula was used [6]:

$$n = \frac{t^2 \cdot p(1-p)}{\Delta_{\omega}^2} = \frac{1,96^2 \cdot 0,28(1-0,28)}{0,05^2} = 309,65$$

There must be a minimum of 310 participants in the sample for it to be representative.

We consider:

t - statistical coefficient corresponding to the probability of guaranteeing the research results

p - the non-percentage share of the sample components possessing the investigated characteristic

Δ_{ω}^2 - standard deviation

We consider $p=0.28$ because in Romania, in 2019, a percentage of 28.1% of the total number of employees worked in companies that are part of multinational groups

controlled from abroad, according to the National Institute of Statistics.

We consider standard deviation to be 5%, therefore $t=1.96$.

The data collection tool

For this descriptive quantitative research, the use of the survey method was chosen. The tool used to collect data within this method is represented by the questionnaire.

Questionnaire:

1. Do you work in a corporation? (Yes/No)
2. What age category do you belong to? (Under 20 years, 20-30 years, 30-40 years, 40-50 years, Over 50 years)
3. On a scale of 1 to 5 (where 1 is not at all familiar and 5 is very familiar), how familiar are you with the field of artificial intelligence?
4. On a scale of 1 to 5, (where 1 is not at all beneficial and 5 is very beneficial), how beneficial do you think artificial intelligence is to your activities?
5. On a scale of 1 to 5 (where 1 is strongly disagree and 5 is strongly agreed), how strongly do you agree with the following statement? "Artificial intelligence will replace a big part of current jobs in the coming years."
6. On a scale of 1 to 5 (1 being strongly disagree and 5 being strongly agree), how strongly do you agree with the following statement? "Artificial intelligence will make my everyday activities easier and easier."
7. On a scale of 1 to 5 (where 1 is strongly disagree and 5 is strongly agree), how strongly do you agree with the following statement? "In the future I will use Artificial Intelligence as a complementary tool for workplace activities."
8. Are you familiar with ChatGPT chatbot launched by OpenAI?
9. Have you used ChatGPT ?
10. On a scale of 1 to 5 (where 1 is not at all and 5 is very often), how often do you use ChatGPT?

11. On a scale of 1 to 5 (where 1 is not at all useful and 5 is very useful), how useful do you think ChatGPT is?
12. On a scale of 1 to 5 (where 1 is not at all concerned and 5 is very concerned), how concerned are you that ChatGPT will negatively affect the educational process?

Research results

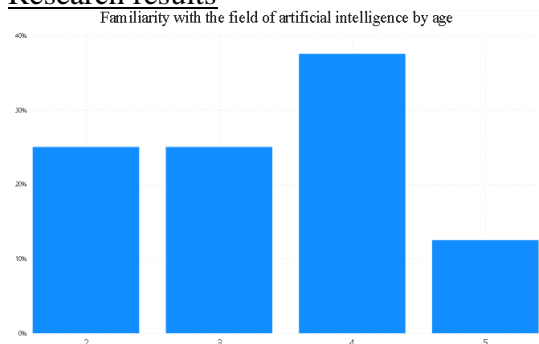


Fig. 1. Familiarity with AI for responses in the under 20 category

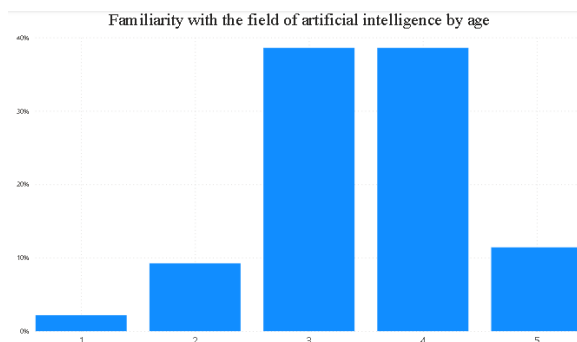


Fig. 2. Familiarity with AI for respondents in the 20-30 years category

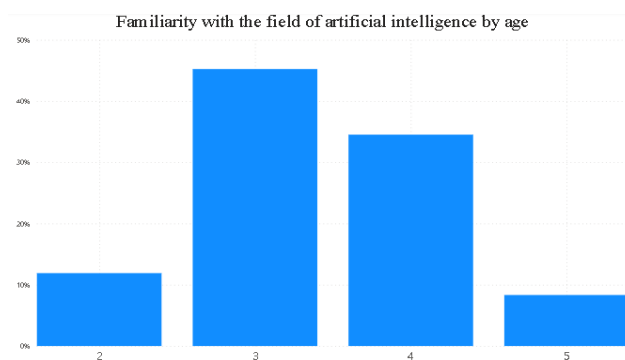


Fig. 3. Familiarity with AI for respondents in the 30-40 years category

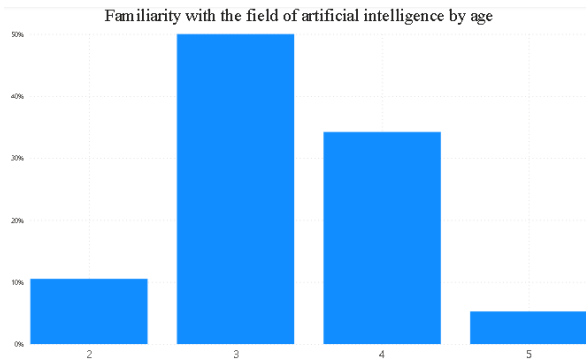


Fig. 4. Familiarity with AI for respondents in the 40-50 years category

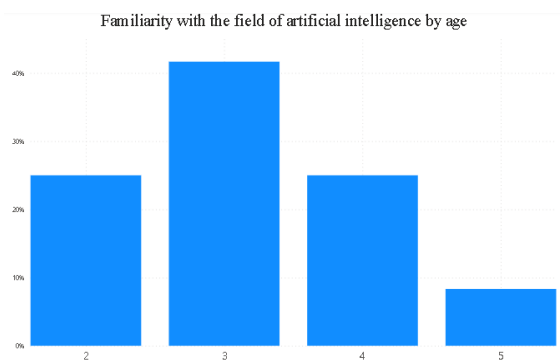


Fig. 5. Familiarity with AI for respondents in the over 50 category

From the graphs presented, it is noted that 37% of young people under the age of 20 say they are familiar with the field of artificial intelligence. Respondents who are in the age category of 20-30 years claim 39% that they are familiar with the field of artificial intelligence. On the other hand, they have a neutral attitude in the same proportion, in addition, 2.17% of people in this category claim that they are not at all familiar with this field. Regarding the age categories 30-40 years and 40 -50 years, they are neutral in the proportion of 45% and 50% respectively. The category of respondents over 50 years old is also 41% neutral.

It can be seen that people in the category under 20 are the most familiar with the field of artificial intelligence. The only category in which a percentage of 2% chose the "not at all familiar" option is the 20-30 year old category. The largest number of people (25%) who chose the option "a little familiar" are in the over 50 age category.

The percentage of respondents who believe that artificial intelligence is beneficial

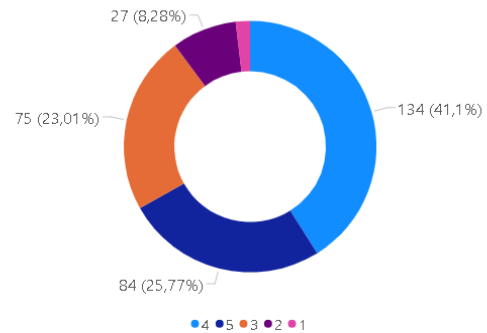


Fig. 6. The percentage of respondents who believe that artificial intelligence is beneficial

It can be seen that a large percentage of respondents believe that artificial intelligence is beneficial (41%) and very beneficial (25.77%).

Replacing current jobs with artificial intelligence

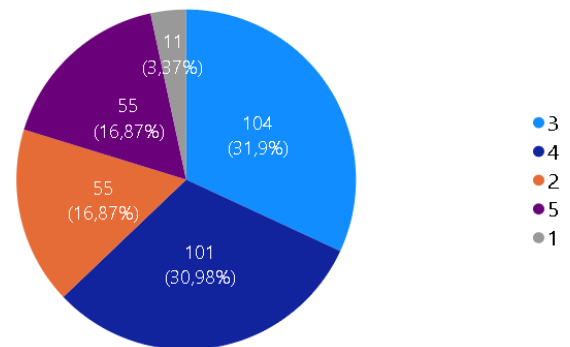


Fig. 7. Replacing current jobs with artificial intelligence

There is a tendency to say that artificial intelligence will take over some of the current jobs. Only 3.37% of respondents disagree with the statement "Artificial intelligence will replace a big part of current jobs in the coming years."

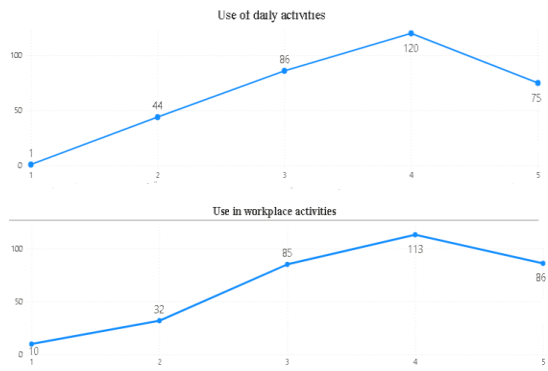


Fig. 8. AI in daily activities and at work

It can be seen that the graph of the efficiency of daily activities using artificial intelligence has the same evolution as the graph of the use of artificial intelligence at work. It can be concluded that the respondents consider artificial intelligence useful both in terms of the workplace and

other daily activities.

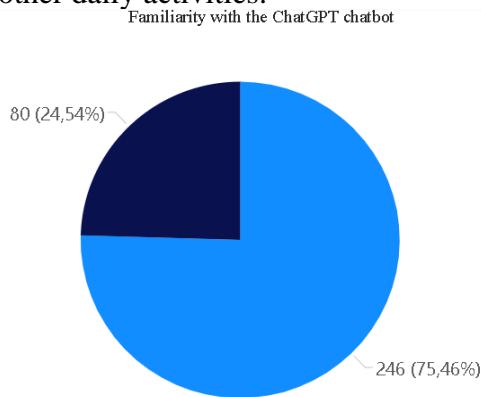


Fig. 9. Familiarity with the ChatGPT chatbot

75.46% of the respondents declare that they are familiar with Chat GPT.

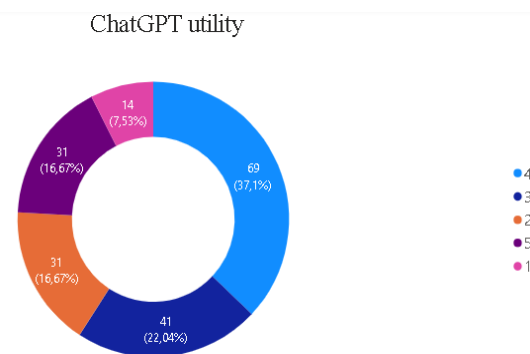


Fig. 10. ChatGPT utility

It can be seen that a percentage of 37% of people who are familiar with the ChatGPT chatbot declare that they find it useful.

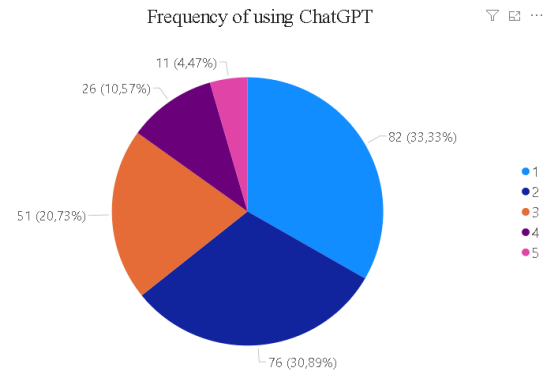


Fig. 11. Frequency of using ChatGPT

It is observed that a percentage of 33% of the respondents say that they did not use Chat GPT frequently.

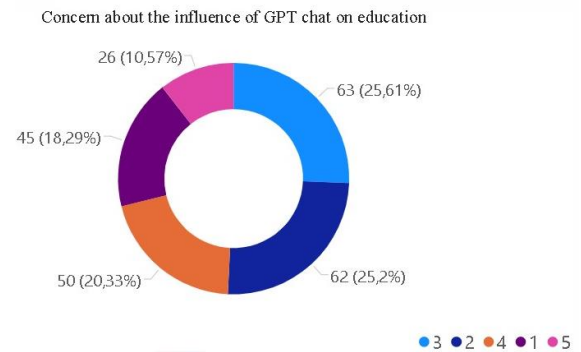


Fig. 12. Concern about the influence of ChatGPT on education

About half of the respondents declare themselves slightly worried or neutral about the influence of Chat GPT in education. On the other hand, 10% of respondents are very worried and 18% are not worried at all about this.

10 Conclusions and future research

Artificial intelligence is a vast field that has been in continuous development over the last century and will be increasingly present in people's lives.

At first mentioned only in fiction books and presented as an abstract subject, it has nowadays materialized in robots and technologies capable of imitating human behavior. The fact that it has applicability in

many fields, it could become a threat to people's jobs. Some countries have already started creating tools to replace human, from self-driving cars to hotel assistance. The popularity of artificial intelligence is a source of stress especially for women and young people.

In addition, its use in education calls into question academic ethics and integrity by facilitating fraudulent answers to assignments and exams. The information transmitted by tools such as ChatGPT is not always completely correct and is for informational purposes only.

In Romania, the population least familiar with artificial intelligence is made up of the over 50s category. Respondents are aware of the benefits of AI, but at the same time they also fear the changes that will occur in the labor market, with 16.87% of the sample believing that artificial intelligence will certainly replace their jobs in the future. Of the people who have heard of ChatGPT, 4.47% already use it very frequently, and 10.57% are very worried about the negative impact it will have on education.

A possible future direction of research in this field could be represented by the study of several categories of the population, to see the impact of artificial intelligence in other fields as well, such as vocational ones.

References

- [1] Anjila, F. (2021). Artificial intelligence. Learning Outcomes of Classroom research, 65-66.
- [2] Apetrii, N. (2022). Inteligența artificială în învățământ. Studia Universitatis Moldaviae, [online]5(155).
- [3] Belanche, Daniel & Casaló Ariño, Luis & Flavian, Carlos & Schepers, Jeroen. (2020). Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success. Journal of Service Management.
- [4] Borana, J. (2016). Applications of Artificial Intelligence & Associated Technologies. [online] www.semanticscholar.org.
- [5] Bleumink, G. and Shikhule, A. (2023). Keeping AI Honest in Education: Identifying GPT-generated text. [online] Available at: https://cdn.aicheatcheck.com/Keeping_AI_Honest_in_Education_Identifying_GPT_generated_text.pdf.
- [6] Cătoiu, I, Bălan, C., Boboc, Ș., Cătoiu, I., Olteanu, V., Pop, N. A. și Teodorescu, N., Cercetări de marketing: tratat, 2009.
- [7] Collins, C., Dennehy, D., Conboy, K. and Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. International Journal of Information Management, [online] 60, p. 102383.
- [8] Cotton, D.R.E., Cotton, P.A. and Shipway, J.R. (2023) Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. Innovations in Education and Teaching International, pp. 1-12.
- [9] Cuturela, A., Gheorghe, A.-M., Simion, C. and Toma, S. (2021). Accesul populației la tehnologia informațiilor și telecomunicațiilor. [online] Institutul Național de Statistică.
- [10] George, A.S., George, A.S.H. and Martin, A.S.G. (2023). A Review of ChatGPT AI's Impact on Several Business Sectors. Partners Universal International Innovation Journal (PUIIJ), [online] 01(01), pp. 9-23.
- [11] Iordache, I. (2021). Revista GeoPolitica. [online] GeoPolitica. Available at: <https://www.geopolitic.ro/2021/11/impactul-inteligentei-artificiale-ia-educatie-si-regandirea-rolului-educatiei-era-digitala/>.
- [12] Jucan, S. (2023). Italia a blocat accesul la ChatGPT / Motivele invocate și avertismentul pentru OpenAI. [online] science.hotnews.ro.
- [13] Kamran, S., Farhat, I., Talha Mahboob, A., Gagandeep Kaur, A., Liton, D.,

- Abdul Ghaffar, K., Rimsha, I., Irum, S. and Afifah, R. (2020). The Impact of Artificial intelligence and Robotics on the Future Employment Opportunities. *Trends in Computer Science and Information Technology*, pp. 050-054.
- [14] Kiefer, T. (2005). Feeling bad: antecedents and consequences of negative emotions in ongoing change. *Journal of Organizational Behavior*, 26(8), pp. 875-897.
- [15] Koo, B., Curtis, C. and Ryan, B. (2020). Examining the impact of artificial intelligence on hotel employees through job insecurity perspectives. *International Journal of Hospitality Management*, [online] 95, p. 102763.
- [16] Masood, R. (2023). Guide to ChatGPT for Parents and Caregivers | Common Sense Media. [online] www.commonsensemedia.org.
- [17] McCarthy, J. (2007). WHAT IS ARTIFICIAL INTELLIGENCE? Computer Science Department Stanford University CA 94305.
- [18] Mokyr, J., Vickers, C. and Ziebarth, N.L. (2015). The History of Technological Anxiety and the Future of Economic Growth: Is This Time Different? *Journal of Economic Perspectives*, [online] 29(3), pp. 31-50.
- [19] Morikawa, M. (2017). Who Are Afraid of Losing Their Jobs to Artificial Intelligence and Robots? Evidence from a Survey. [online] www.econstor.eu.
- [20] Muthumanicam, S.K., 2018. The Evolution of Artificial Intelligence (AI) in Hotel Industry. August. Retrieved from: <https://chatbotnewsdaily.com/the-evolution-of-artificial-intelligence-ai-in-hotel-industry-23298d1e8dd4>.
- [21] Nam, T. (2019). Technology usage, expected job sustainability, and perceived job insecurity. *Technological Forecasting and Social Change*, [online] 138(C), pp.155–165.
- [22] Newell, A. and Simon, H.A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), pp. 113-126.
- [23] Rampersad, G. (2020). Robot will take your job: Innovation for an era of artificial intelligence. *Journal of Business Research*, 116, pp. 68-74.
- [24] Rathore, D.B. (2023). Future of AI & Generation Alpha: ChatGPT beyond Boundaries. *Eduzone: International Peer Reviewed/Refereed Multidisciplinary Journal*, [online] 12(1), pp. 63-68.
- [25] Sagara, H. and Das, K. (2019). Technological Disruptions and the Indian IT Industry: Employment Concerns and Beyond. *Digitalisation and Development*, pp. 119-143.
- [26] Shahriar, Sakib & Hayawi, Kadhim. (2023). Let's have a chat! A Conversation with ChatGPT: Technology, Applications, and Limitations.
- [27] Skilton, M. and Hovsepian, F. (2018). *The 4th Industrial Revolution*. [online] Cham: Springer International Publishing.
- [28] Smith, S. (2017). StackPath. [online] www.ehstoday.com.
- [29] Verheyen, S. (2021). RAPORT referitor la inteligența artificială în educație, cultură și sectorul audiovizual [A9-0127/2021] Parlamentul European. [online] www.europarl.europa.eu.
- [30] Westfall, C. (2023). Educators Battle Plagiarism As 89% Of Students Admit to Using OpenAI's ChatGPT For Homework. [online] Forbes. Available at: <https://www.forbes.com/sites/chriswestfall/2023/01/28/educators-battle-plagiarism-as-89-of-students-admit-to-using-open-ais-chatgpt-for-homework>



Elena COSTIN is a graduate of the Faculty of Economic Cybernetics, Statistics and Informatics at the Bucharest University of Economic Studies Economic Informatics, with a Bachelor's degree in Economic Cybernetics. She has recently completed a Master's degree in Databases - Support for Business and has a keen interest in Big Data, Business Intelligence and Data Analysis.



Andreea Georgiana ENE is a graduate the Faculty of Economic Cybernetics, Statistics and Informatics at the Bucharest University of Economic Studies, with a Bachelor's degree in Economic Informatics. She continued down the path of Economic Informatics and she recently graduated a Masters' degree in Databases-Support for Business. She is passionate about the financial - banking field, web design and the legal field. She likes art, she likes to color applications, the financial field because it teaches her to be financially and legally organized.



Elena Daniela DUMITRAȘCU has successfully obtained her Bachelor's degree in Economic Sciences and Business Administration from Brasov, specializing in Economic Informatics. She further pursued her academic journey by enrolling in the master's program offered by the Faculty of Cybernetics, Statistics, and Economic Informatics, where she completed the Business Support Databases curriculum. Among her diverse range of interests, programming web applications stands out as one of her notable passions.



Florinela Georgiana PREDA is a graduate of the Faculty of Economic Cybernetics, Statistics and Informatics at the Bucharest University of Economic Studies, with a Bachelor's degree in Economic Cybernetics. She has recently achieved a Master's degree specializing in Databases - Support for Business and her domains of interest are Data warehousing, Data Analysis and Data science.

Insights and proposals for RPA implementations

Andreea Izabela BOSTAN¹ and Ana Ramona BOLOGA²

^{1,2} Bucharest University of Economic Studies, Bucharest, Romania

izabela.bostan@csie.ase.ro, ramona.bologa@ie.ase.ro

Abstract: *This article presents a comprehensive bibliometric analysis of research articles focused on Robotic Process Automation (RPA) project management. By analyzing a large dataset of scientific papers, this study aims to identify trends and gaps of this subject. To better understand the real-life implications, this study also analyzes different opinions coming from people actively working with this technology. The bibliometric analysis aims at identifying several key themes often addressed in scientific papers and correlations, while also understanding the interest in this field among re-searchers. The analysis is based on a correspondence between Intelligent Automation concepts and methodologies for projects' development. The analysis continues with the results of a survey completed by a more detailed series of interviews at the beginning of 2023 that focuses on a real-life perspective, with the objective to identify project phases where teams are often experiencing challenges.*

Implementation of Robotic Process Automation initiatives depends heavily on the project lifecycles, however RPA in the context of project methodologies is a topic not sufficiently researched at the moment. RPA teams are expressing different preferences regarding implementations, Agile developments being one of them, as it seems that Agile principles are closely matching RPA criteria. Furthermore, based on the findings, this article proposes a set of practical suggestions to enhance the success of RPA implementations in different project phases. The originality of this paper is reflected in the methodology adopted, that includes different techniques, in the attempt to complete each other: the literature overview and current perspectives.

Keywords: *RPA, Intelligent Automation, Lightweight IT, software lifecycles, Agile*

1 Introduction

Robotic Process Automation (RPA) is a technology that uses software robots to automate routine, rule-based actions in corporate processes. RPA bots execute tasks including data entry, data extraction, form filling, data manipulation, and more by simulating human interactions with digital systems and applications. RPA bots are designed to interact with user interfaces of existing software applications, just like a human user would. They can navigate through different screens, input data, retrieve information, perform calculations, and execute predefined actions. RPA technology does not require changes to the underlying systems or applications; instead, it leverages the existing user interface to interact with them.

RPA is one of the most advanced technologies of the moment that can

change the way business processes are carried out within organizations [1] emerging as a vital and strategic catalyst for sustainability [2]. As a result of the situation generated by the COVID-19 pandemic, many companies have switched to business workflow automation, which has accelerated the increase in demand in this field. According to some market studies [3,4], the global RPA market size was valued at 2329.9 million dollars in 2022, and is expected to rise at a CAGR of 35% from 2022 to 2030.

Organizations use RPA to automate manual and repetitive tasks across various industries and departments. It helps improve operational efficiency, reduce errors, and free up human workers to focus on higher-value activities. RPA can be particularly beneficial in areas such as data entry, data migration, invoice processing, customer service, finance and accounting, HR

processes, and more. This type of software robots can be scheduled to run at specific times or triggered by events. At the same time, RPA platforms often provide monitoring and analytics capabilities to track the performance and effectiveness of the automated processes. RPA is different from traditional software development [5], as it focuses on automating tasks within existing systems without requiring significant changes in the underlying infrastructure or applications. It provides a quick and flexible solution for automating repetitive processes, allowing organizations to achieve efficiency gains and cost savings. Considerations such as process complexity, scalability, and ongoing maintenance should be carefully evaluated to ensure successful implementation and long-term sustainability. Overall, the sustainability of RPA depends on how it is implemented, managed, and integrated into an organization's broader sustainability strategy. When implemented responsibly, RPA can contribute positively to environmental, economic, and social aspects of sustainability [6]. By automating tasks that typically involve printing or handling physical documents, RPA can help organizations reduce their need for extra printouts and promote eco-friendly practices that are saving time and energy. As a result of RPA implementation, resources will be redistributed and strategic priorities could take precedence, as employees are refocusing their allocation on higher value activities. As RPA benefits and sustainability have been increasingly emphasized and noticed in more and more industries and RPA started to gain a non-negligible importance, today it has reached a point which places it as a technology with a huge potential for future implementations [3,4]. The maturity of the technology itself, as well as the future perspectives for RPA developments are automatically

setting the need of a methodology that will define a set of principles and procedures, with the aim to contribute to the optimum performance of the software robot implementation.

To ensure adequate coordination, resource allocation, risk management, and on-time delivery throughout the RPA project lifecycle, effective project management and governance are crucial. Project lifecycles in IT developments offer a structured framework for organizing, carrying out, and managing the complete project from initiation to closure. Adopting a project methodology facilitates collaboration between business users, RPA developers, and other stakeholders, resulting in a seamless and effective automation implementation. Dividing the workload into project lifecycles and adopting specific ideologies also enables businesses to improve continuously, by learning from each project iteration and applying the best practices and lessons gained to new RPA efforts.

Our study has significant value for research and practice in IT, RPA and RPA project management because it provides a comprehensive synthesis of the existing literature on the relationship between RPA and software implementation methodologies. This paper is based on the following research question: "What life cycles does an RPA project follow and what are the main specifications for these?", and aims to identify these concepts in the specialty literature on the one hand, and based on the specialists' experience in practice, on the other hand.

To start with, in order to first get a better understanding of the topic, a literature review is developed, to identify to what extent this subject has been addressed by researchers. In order to achieve the same objective, a bibliometric analysis is performed. Moreover, the methodology is completed with an overview of RPA implementations nowadays, described by analyzing a set of data gathered after conducting a questionnaire and a series of

interviews. These approaches are meant to identify: what are the main challenges that RPA teams are facing, during what project phases are challenges identified, and what best practices could be followed in order to minimize the number of these. Finally, the study proposes a set of suggestions, gathered during the survey, interviews and by analyzing the current literature, with the aim to propose a set of recommendations that can minimize challenges and contribute to the success of an RPA implementation.

2. Foundations of RPA implementation process

UiPath [7] states that “the emergence of the term “robotic process automation” can be dated to the early 2000”, even if “the origins of the term “workflow automation” dates back to the 1920s during the industrial era and emergence of manufacturing” [7].

Gartner defines RPA as “a tool that performs “if, then, else” statements on structured data, typically using a combination of user interface interactions, or by connecting to APIs to drive client servers, mainframes or HTML code” [8]. The benefits generated by the use of RPA are numerous and have sparked a wave of enthusiasm and many followers. Some of the advantages that RPA contributes to are: ensuring that corporate procedures and operations adhere to rules and compliance standards, drastically boosting the rate of processing, and increasing effectiveness through process data auditing and digitization [9-10]. But there is a lack in the area of theoretical re-search and conceptual frameworks [11].

An important number of RPA projects fail, generating concern for studying the success factors and the challenges encountered in their implementation [12]. The skepticism of many comes from the idea that the use of robots will lead to job losses and massive layoffs, which has not

happened until now. In addition to this, the misunderstanding of RPA features, compatibility issues with IT infrastructure, skill sets and security protocols, are elements that affect and hinder the implementation of RPA in companies [13]. When discussing implementation of IT projects, one topic that is essential to the subject is how the development is going to be carried out, what is the methodology that the team is going to adapt to and what are the main phases that will ensure the project success. Different RPA vendors and researchers [14-16] are promoting Agile as the best methodology that should be followed in order to achieve most of the automation potential. The definition provided by Atlassian (<https://www.atlassian.com/agile>) is that “the Agile methodology is a project management approach that involves breaking the project into phases and emphasizes continuous collaboration and improvement. Teams follow a cycle of planning, executing, and evaluating”. In a few words, the reason why Agile is promoted among RPA teams is that several principles of Agile Scrum can provide a better framework to deliver RPA software. First of all, RPA aims to re-place the work performed by human agents, so an RPA automated solution must replicate the human component as much as possible. Since the collaboration with the end-user during the implementation is preferable, Agile can promote this approach as it focuses on the collaboration between team members, stakeholders being included as well. Also, it is considered that RPA “does not respond well to change” [16], and the Agile approach is definitely covering this weakness, as it is based on iterative implementations, always leaving room for improvements and prioritizing feedback. Agile also offers a strategy that encourages early conceptualization and design to ensure flow optimization. With requirements being detailed in an early stage, the RPA robot is less likely to suffer major changes in the logic. Moreover, the option of a Backlog

gives the RPA teams the possibility to take into consideration the remaining work and split the effort into multiple iterations, while Agile specific meetings (Planning, Retrospectives) are giving the teams the option to reprioritize the remaining work and better function during the following Sprints.

A traditional software development is characterized by manager-led teams that are organized in a hierarchical structure and have multiple layers of authority [17]. In contrast, Agile teams are designed to be democratic teams, where all members are treated as peers on an equal level and there is no formal hierarchy. This comes as a result of the Agile methodology that focuses on social interactions and on people collaboration, this being one of the Agile manifesto values: “Individuals and interactions over processes and tools” [18]. For an Agile team to work, there are only three roles that should be covered: Product Owner, Scrum Master and the Development team [19]. Schwaber defines the Product Owner as the person “responsible for representing the interests of everyone with a stake in the project and its resulting system” [20]. The Product Owner is responsible for the funding of the project, return on investment (ROI) objectives and also “ensures the most valuable functionality is produced first and built upon”. A Scrum Master acts as a “servant leader”, as the Agile literature de-fines this role [19]. It is a key player as it coaches the team members, helps remove the impediments and ensures all Scrum events take place. The Development Team includes several different roles, such as: Developers, Solution Architect, Business Analyst, Tester, but all are collectively responsible for the project as a whole. Solution Architects design the overall technical vision, while guiding Developers who are involved in delivering the technical solution. The Business Analyst clearly defines and communicates the

requirements, acting as an interface between departments. Agile practitioners consider the team to develop by passing through four stages: forming, storming, norming and performing, this being a cycle that repeats every time a change is introduced [21]. Even if, in practice, companies and teams in general are having different approaches in terms of organizing the development workload, generally speaking, carrying out the software process typically follows six main phases: project planning, analysis, design, development, testing, and roll-out [22].

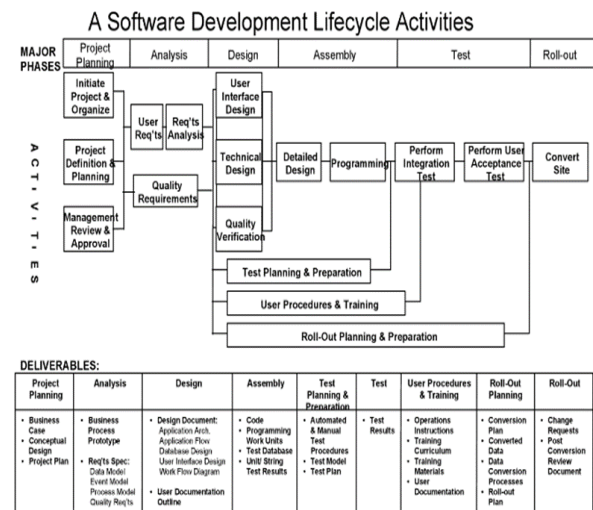


Fig. 1. Software development life cycle activities and deliverables [22]

To describe a general approach of life cycles, the process begins with the Project Planning phase, which involves the identification of Information Technology requirements resulting from business objectives. The next step is the development of an action plan or work breakdown structure to specify specific initiatives to achieve the defined goals. This phase determines the conceptual design's development strategy, advantages, and an estimate of development expenses. In order to complete this phase, it is essential to: initiate and organize the project, create a definition, project scope and a planning, as well as seek management review and approval. The business case's result, which was generated during the project planning phase, provides the justification for the

system and project decisions, and becomes the primary set of business objectives used to govern and direct the project. The relative importance of competencies shifts throughout a project's closing phase, and "negotiation" and "managing stake-holders" are now among the top two competencies [23].

The Analysis phase focuses on gathering and documenting functional requirements that are researched in detail. The degree of automation and the complexity of the process are established. As a result, the advantages and results of automation are immediately highlighted. It is also the stage when the "Solution Design Document" or "Functional Specification Document" is created, based on the previous requirements analysis. This document comes as a detailed definition of the business process that will later be confirmed and approved by the business representatives (the team requesting the implementation). As main topics addressed as part of this documentation, the following items are defined: a description of the entire process, detailing each step that will be implemented in the development, technical details, business requirements, process exceptions, elements identified as out of scope. Researchers [23] are considering this stage to be marked by three main competences: "managing stakeholders", "communicating" and "planning": "Here the project managers build lines of communication and facilitate communication with the client and other stakeholders to capture and create a clear and comprehensive requirements statement".

The Design phase is essential in order to create solid workflows that incorporate all technical constraints and, at the same time, respect the business requirements previously agreed on. In order to create solid RPA workflows that will be later implemented by RPA developers, it is imperative that the process is described in depth, with as many potential exceptions

(both technical and business) understood as feasible. These specifics may include screen designs, databases, data transfer flows, system interfaces, and prototypes, depending on the project.

As part of the Development phase, teams are implementing the technical solution, with reference to the previously created Solution Design Document and all use cases and business rules confirmed. The process is divided into smaller, sequential or parallel steps, so that the product management is improved and the workload fractionated correspondingly.

The Testing stage is essentially divided into different other phases. Not only different environments should be tested, but also the approach would differ from one stage to another, from testing functionalities individually, to observing an end to end flow. In any case, this stage is performed after the implementation of all process flows, based on test data according to those used in reality and previously defined test scenarios. The solution is tested in a pre-production environment to confirm that it meets the process requirements. If defects are discovered at this point, the development phase is restarted to fix all existing issues, an approach that could potentially characterize this as a cyclic flow. This phase is completed with the User Acceptance Testing, a testing phase conducted by the business representatives, initiated with predefined data and seeking approval from the client.

Last but not least, the last stage of a project, known as the "Go Live" phase, marks its completion with the Deployment phase. This is the state when all components are integrated and verified as a whole in the production environment, with a high rate of involvement from the client/ stakeholder side. At this point, the previous work is moved to the actual Production system, and the anticipated advantages, such as cost savings and improved process quality, are tested in real-world operations. After the deployment in the Production environment, it is essential to establish an intensive period

of Hypercare, when the development team will monitor the proper execution of the robot. If a defect is discovered in the application, the team is ready to provide on-going support and maintenance for minor bugs. Studies have determined that this phase registers most of the costs involved with correcting potential errors that may appear in a software system [24]: “Finding and fixing a software problem after delivery can be upwards of 100 times more expensive than finding it and fixing it during the requirements and early design phases”.

Approximately 98% of the implementation of an RPA project is related to business rules, so business process experts play an essential role in implementation. As a result, RPA implementation cannot be equated with classical software implementation [25]. Within the IT sector, Robotic Process Automation is known to be a lightweight IT technology [26]. Lightweight IT is the new paradigm of mobile apps, sensors and bring-your-own-device, also called consumerization or Internet-of Things. The key aspect of lightweight IT is not only the cheap and available technology as such, but the fact that its deployment is frequently done by users or vendors, bypassing the IT departments ([27]). To better understand the concept of “lightweight IT”, it is also essential to mention the definition of “Heavyweight IT”, which, in comparison, is the mainstream IT as currently delivered by IT departments over the world: back-end solutions such as ERP and other transaction systems, based on databases servers and integration software, such as bus architectures [28].

Table 1. Heavyweight and lightweight IT [28]

	Heavyweight IT	Lightweight IT
Profile	Back-end: Supporting documentation of work	Front-end: Supporting work processes

	Heavyweight IT	Lightweight IT
Systems	Transaction systems	Process support, apps, BI
Technology	Servers, databases, enterprise bus technology	Tablets, electronic whiteboards, mobile phones
IT architecture	Centralized or distributed	Meshworks
Owner	IT department	Users and vendors
Development culture	Systematics, quality, security	Innovation, experimentation
Problems	Increasing complexity, rising costs	Isolated gadgets, security
Discourse	Software engineering	Business innovation

As described in Table 1, even if the difference between these two concepts are obvious, they can be considered complementary and mutually dependent. Lightweight IT may not have the full control on data repositories and platforms, as Heavyweight IT does, but it brings the innovation and agility that can be considered as out of scope for the Heavyweight IT.

That being said, RPA can be easily framed as a Lightweight IT as it simply inter-acts with IT systems’ presentation layer in order to simulate human behavior. Fundamentally, RPA is a technology that connects all other IT applications non-invasively and seamlessly. As such, as Moayed states in the “From pilot to full scale RPA deployment” whitepaper [29], it is possible for RPA solutions to become the “platform” or “backbone” through which all other automation technologies will eventually be connected/ organized, be it Chatbots, Optical Character Recognition or simply other soft-ware platforms. Even if RPA classifies as a component of the IT sector, and the delimitation between projects’ phases can be applied as for any other software project, the distinction between Lightweight and Heavyweight is one reason which needs to be considered when rethinking implementation phases in the RPA context. The organization’s ongoing business and IT alignment, as well as the

overall limited utilization of RPA in an organization [30] are factors that could add different specifications, compared to the traditional Heavyweight IT approach.

3. Methodology

The current study is based on two different methodologies that were adopted in order to better understand the subject and develop several conclusions and recommendations: a bibliometric analysis and a questionnaire conducted among RPA companies.

First of all, a bibliometric analysis was carried out in order to gain insight into current research trends and the extent of study in certain research topics. This particular statistical analysis of publications tracks researchers' output and impact and publication relationships can also be understood by visualizing the bibliometric data. It is a remarkable tool for the measurement of academic and organizational performances based on various indicators such as the number of publications and citations, collaboration networks and examines the research productivity, most keywords and publication trends in a particular research area [31]. It also allows easy identification of re-search gaps in a particular research field.

In terms of data collection, the study was conducted by utilizing data from the Scopus database, in order to identify a set of bibliometric indicators, such as quantitative indicators and qualitative indicators that specify citations. Scopus is a multidisciplinary citation database of peer-reviewed literature with tools to track, analyze, and visualize research [31] and so it is considered to be a reliable source that offers accurate information. Identifying various research-related publications' sources and recognizing trends in annual publications were the main objectives considered when applying this type of research. A total of 268 publications were analyzed as part of

the data collection with the use of Scopus database, having a timeline, from 2009 to 2023. The data was filtered by specific criteria (detailed in the next chapter) and then ex-ported in a Comma-Separated Values (CSV) file. The tool used to perform the analyses on the exported data was VOS Viewer [32], chosen in order to perform the bibliometric analysis, with outputs in the form of visual components based on mapping techniques. The data stored in the CSV file was translated into clusters and diagrams, with the aim to assess relevant information about the publications.

Furthermore, in order to provide a clear statistical analysis regarding RPA implementations today, the study uses the results gathered as part of a questionnaire addressed in February 2023, with the aim of getting a better understanding of how RPA tools are used to develop different projects, in order to identify how this technology is currently affecting internal business processes and it is expected to evolve in the near future.

The questionnaire was addressed to 150 respondents that are closely working with the RPA technology and received a total of 122 unique responses, having an 81,33% response rate. The 122 respondents are using RPA in 15 different industries. All 122 responses have been validated, as the criteria based on which an individual was chosen as a potential respondent was to have been working with RPA at least for the last two years, regardless of the company or role.

The questionnaire was created using Google Forms and has been shared online, mostly by email and on a professional social media platform. Results were later interpreted by displaying the data in a visual format such as pie charts and bar charts. While bar charts only highlight the number of responses and differences between answers are easier to remark, pie charts represent each result as a percentage of the whole sample.

The complete study was based on a total of 22 questions, divided into three sections: the profile of the respondents, a theoretical approach of the RPA technology on a

conceptual level and applying these concepts in real life projects using different RPA vendors. However, for the objective of the current paper, only six of the total questions were taken into consideration and analyzed, in order to understand possible draw-backs of the technology that could reside in potential areas of improvement. A more detailed analysis has been conducted as part of a different research paper [33].

Responses gathering did not imply any personal data collection as the profile of the respondents was built as part of the first section of the questionnaire. This one has been defined by questions analyzing the following topics: the industry to which their company belongs, the department they are working in, their location, experience with the RPA technology, role and the number of RPA projects they have experienced in the last 12 months. All this information was later correlated with the results observed on the other questions. All questions were mandatory, with multiple choices but single answer and the option to add a different response, if needed for the company industry, department, or role. Respondents are mainly located in Romania (95%) and have been working with RPA technologies for different periods of time: 2-5 years (59%), less than 2 years (29%) or more than 5 years (12%).

Having the survey, which aims to understand the way RPA teams work nowadays, as a starting point, a series of more in-depth interviews have been conducted in the same time period. The objective of this interview series was to get a better understanding of some of the topics raised by the questionnaire. The “interview” method has been chosen as an appropriate tool to collect detailed and contextual information from individuals. As Barbour mentioned, the interview aims at having an ‘in-depth information’ about a certain topic or subject, and through which a phenomenon could be

interpreted in terms of the meaning’s interviewees bring to it [34]. This way, the inter-views intended to allow some of the questionnaire’s respondents to provide subjective experiences insights, while also providing the opportunity to address follow-up questions, to better understand participants’ viewpoints and resolve ambiguities.

4. Results and discussions

The Bibliometric analysis was developed based on complete data collection and data filtering procedure. To begin with, the primary goal of the research was set to analyze the trends in publications linked to Robotic Process Automation and methods of Project Management, in the Scopus database. The research was done utilizing the document examination method within the context of descriptive analysis. The data was retrieved on May 7th, 2023, using “Robotic Process Automation” and “methodology” as main keywords and “phases” and “implementation” as secondary keywords. A total of 268 documents were retrieved following this procedure, these being publications that are addressing these defined topics.

The inclusion criteria were the document type and language, in order to make sure the publications complied with academic norms. Results were filtered to exclude non-English publications and the following document types: erratum, retracted papers, conference review. Only the following have been analyzed: articles, conference papers, book chapters, books and reviews, which led to a total of 251 publications, after applying the initial filters. The analysis showed that the majority of documents were journal articles (a total of 117), as well as a remarkable number of papers published in different Conferences (97 out of the 251 total).

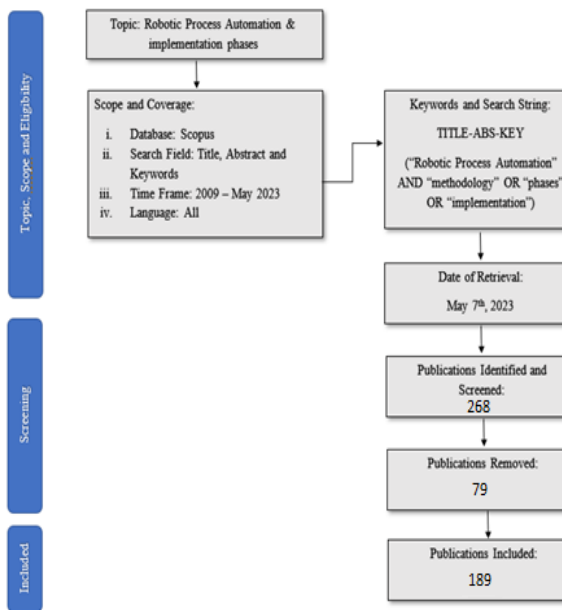


Fig. 2. Selection strategy and research protocol

In order to ensure the relevance of the results, a manual review process was conducted. During this phase, the title, abstract and keywords of each research publication were evaluated for potential relevance to the study, to validate that the paper's content is developed based on the set of keywords provided in the search criteria. As a result, a number of 62 documents have been excluded from the data set, as the abstract proved to be irrelevant for the context of analyzing Robotic Process Automation implementations.

After this additional filter, the timeline considered has been affected as well. For example, without excluding any papers, the initial unfiltered results show that the first article that follows the direction input in the search query was published in 2009. However, by analyzing the abstract it is clear that the article focuses on robotics and process automation as hardware configurations and a methodology is suggested to design only these kinds of systems. With that being observed, the final timeline addressed turned out to be 2017 - 2023.

It is interesting to see how the interest in

this topic developed during these five years. With first initiatives on analyzing the implementations of RPA technologies in 2017 and only two relevant documents retrieved by the search for that year, the interest has drastically increased during the next few years, with a maximum of 57 publications in 2021. In this rapid evolution, it can be assumed that the topic will continue to gain interest around researchers for the next year as well, considering that, until this moment, there is a number of 31 documents published until May 2023. However, important to mention is the fact that these numbers are not reflecting only publications that follow the initial filtering based on all keywords exactly. Most of these results have as a central point the concept of "Robotic Process Automation" itself, and describe a specific implementation, use case in an industry or highlight a specific RPA concept indeed. In spite of that, they are not suggesting a specific methodology or framework that maximizes the implementation of RPA processes by following a project methodology. Moreover, the concept of "RPA" in the context of methodologies or implementation phases is addressed in just a few publications from this set of results, and it details specific behaviors of RPA implementations (i.e.: Jimenez-Ramirez, A. et al., 2019, A method to improve the early stages of the robotic process automation lifecycle).

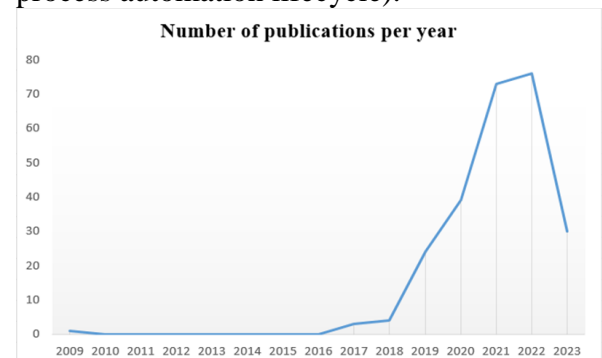


Fig. 3. Number of publications per year, unfiltered results (a total of 268 publications)

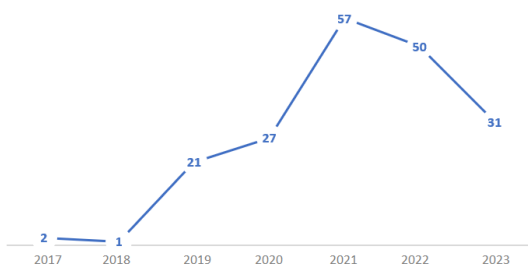


Fig. 4. Number of publications per year, results filtered by document type, language and abstract relevance (a total of 189 publications)

As the tool (VOS Viewer) used to create data visualization representations is creating analysis based on identifying the most used terminology (keywords) and the number of occurrences, for the co-occurrence analysis the tool automatically identified a number of 1465 keywords that were later filtered to only cover the RPA technology and any possible relation with implementation phases or methodologies, as it can be observed in Fig. 5. As a result, the all keywords co-occurrence visualization shows a network diagram of keywords, in which nodes, font size and color are describing the relationship with one another. Fig. 6 only shows a selection of this network, where one of the most popular keyword (“rpa”) is grouped into a cluster with other keywords with a medium occurrence, such as: “business process”, “enterprise resource management”, “business process management” or “process analysis”. The membership to the same cluster is represented by the same color used for the connecting lines, and grouping these concepts into one cluster demonstrates that RPA is closely related to the idea of improving business processes and managing projects. However, no clear evidence could be observed between RPA and specific project methodologies concepts, concluding that this is still a topic not sufficiently researched.

Keyword	Occurrences	Total link strength
robotics	89	1161
automation	42	519
rpa	47	445
artificial intelligence	32	372
business process	27	352
intelligent robots	19	292
enterprise resource management	19	269
digital transformation	20	245
robotic process automation (rpa)	21	195
business process management	17	169
data mining	11	166
decision making	10	163
machine learning	15	151
information systems	12	149
intelligent automation	11	149
process mining	14	144
information use	10	140
botnet	9	136
life cycle	8	123

Fig. 5. List of keywords filtered for the bibliometric analysis

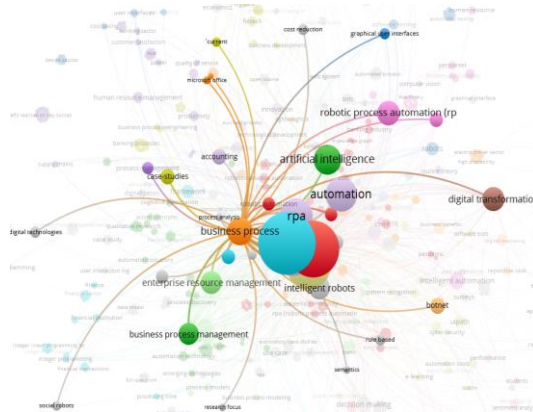


Fig. 6. All keywords co-occurrence

Fig. 7 is a representation of a co-authorship analysis by countries, in a density visualization, showing how the maximum research around this topic took place in: Germany, India and the United States.

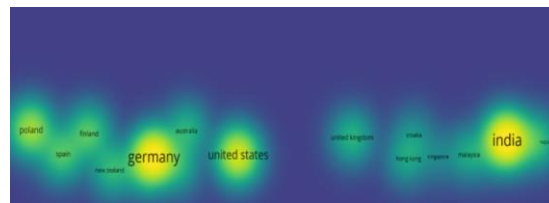


Fig. 7. Main countries associated with publications

While the bibliometric analysis highlighted the necessity of deepening this specific topic as it couldn’t provide a clear input on the research question stated in the beginning, the questionnaire conducted managed to identify actual business needs and get an understanding of the current approaches when it comes to RPA development. One topic addressed as part of the questionnaire conducted was related to the usual RPA

project lifecycle, with a focus on the project phases in which teams are often facing difficulties. A majority of 30% of respondents are considering the Analysis phase to be the most error prone. Design and Development phases are the next preferred options, with a 20% choice rate each. By analyzing these numbers, it is also important to mention that most of the respondents are working as a Developer/Architect (74%), the difference between the other existing roles being remarkable: Business Analyst/Product Owner (13%), Management positions (8%), Project Manager/Scrum Master (2%). In this context it is interesting to notice to what extent the initial phases of a project (Discovery, Design) are considered to be the most challenging in terms of the number of risks that could potentially affect the whole project implementation. Going even more into this subject, the fact that the initial phases of a project are usually the most demanding is confirmed by responses received to a different question: “What are the challenges that may be experienced during a project development?”.

Considering the usual RPA project lifecycle, what is the phase in which the team is most often facing challenges?

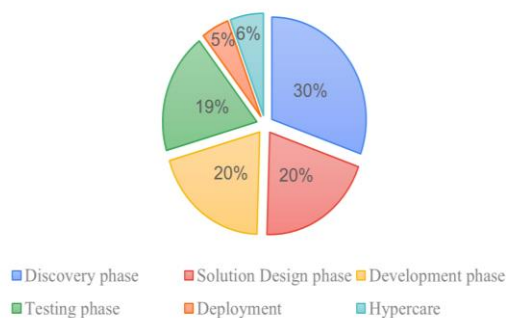


Fig. 8. Project phases where teams are having most of the difficulties

Respondents were allowed to select multiple choices, and the results show that the top three most difficult aspects are faced when: the initial manual process is not clear enough or is lacking adequate

documentation (with a reference to the first phases of the project implementation), customers are expressing reluctance or adding complexity to the implementation during the process and when edge cases are not covered by the software robot (which, again, can be referenced to the same Analyze and Design phases).

Among responses, 25,71% of the results are targeting the adoption of an unsuited project methodology. The “Pulse of the Profession” study conducted by the Project Management Institute (PMI) in 2018 concludes that 71% of organizations were choosing the Agile approach, even if applied to different methods, based on each team's needs. This number was a result of a survey that highlighted feedback from 5702 Project Managers and other management roles, from a range of industries, including In-formation Technology. “More and more organizations are recognizing that agility - the capability to quickly sense and adapt to external and internal changes to deliver relevant results in a productive and cost-effective manner - is helping them stay competitive” [35].

The same conclusion can be highlighted considering the study that made the subject of this paper, as 72% of respondents considered that Agile methodologies can best be associated with automation processes.

What are the challenges that may be experienced during a project development?

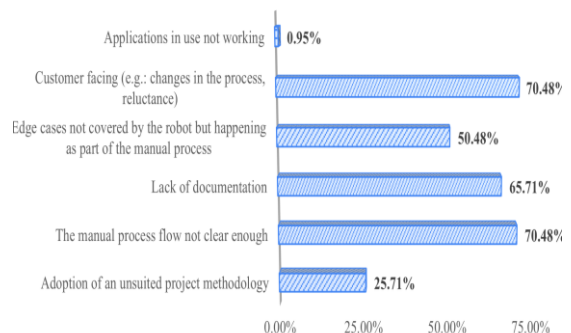


Fig. 9. Challenges experienced during an RPA project development

Having as a starting point the topics addressed as part of the questionnaire, the most relevant responses provided have been

detailed by some of the participants as part of a more comprehensive interview. Participants have been asked “What are the best practices that teams are adopting during each project lifecycle, when automating processes using RPA?”, in order to understand what particularities can be considered when discussing Project Management for RPA driven projects. By analyzing the general project life cycle of a software implementation, as detailed in the Literature Review section, and by understanding challenges identified in RPA projects development in real life implementations, Table 2 describes a set of recommendations, with the objective of capturing a series of additional requirements that RPA projects have, compared to a usual IT process. The following ideas and recommendations can be viewed as a theoretical framework for RPA implementation, considering the generality of each action.

Table 2. RPA projects’ requirements for each implementation phase, in addition to general software development need

Project Phase	RPA projects’ requirements, in addition to general software development needs
Project Planning	Highlight the weaknesses of a manual process and propose a new, more efficient way of achieving the same or even better results. Anticipate use cases that are occurring in day-to-day business, but may not be obvious from the beginning. If scenarios are not considered from the planning phase, then the RPA robot’s configuration will have to be adjusted or the scope of the automation would have to be expanded, in order to include these unknown cases. This results in affecting the project’s timeline, resources and costs, not to mention the customer satisfaction.

Project Phase	RPA projects’ requirements, in addition to general software development needs
Analysis	To better understand the process flow, develop process diagrams: a diagram that will include the steps described in the initial non-automated flow and another diagram of the automated process.
Design	Create an architecture that follows a tactical approach and consider using RPA benefits at its full extent. For example, if none of the processes chosen for the automation are not performed at night, having an RPA robot working 24/7 may not have a visible impact.
Development	An RPA software selection process is mandatory: the selection of appropriate RPA software for automation is the main emphasis of this step. Fortunately, the market appears to be maturing swiftly, and so some essential considerations in the decision-making process include: the cost of the vendor, required skills, vendor support, vendor reputation, the capability to use low-code programming, security, license flexibility and so on.
Testing	The test environment should be an exact replica of the Production system in order to ensure efficiency and avoid rework in configuring the robots. A robot might fail to recognize a field in the production system if it differs from the one in the test environment, throwing an error that would ultimately interrupt the process. RPA robots are dependent on User Interface components that are set as triggers when creating the logic behind the automated process, so it is essential that these components that are used initially in the development are matching exactly the real ones.
Roll-out	Anticipate customers’ resilience when it comes to being open to adapt to RPA automated

Project Phase	RPA projects' requirements, in addition to general software development needs
	<p>technologies, which is also confirmed by the questionnaire conducted for this study.</p> <p>Regarding the customer relationship involved in RPA processes, 76,19% of people that are often facing the end user's reluctance are connecting this with the lack of knowledge that the customer may have towards this technology. They also reported the idea that RPA robots could take over their tasks, or they don't believe in minimizing or excluding manual work. It is important that the end user understands from the beginning the logic behind the implementation, RPA implications, and see the concrete results and benefits of the automation. It is also important for the end user to understand the limitations and have a clear image of what are the out-of-scope elements, so that expectations are realistically set.</p> <p>Consider an Agile approach for moving robots to Production, in order to provide efficient deployment and close monitoring of processes. As a best practice it would be recommended to transfer to the Production environment a smaller number of robots to carry out (a part of) a process, identify and implement the potential adjustments/ bug fixes, then transfer into production other RPA robots, until the target system is accomplished.</p>

5. Conclusions

With the concrete classification of RPA as part of the IT segment and the theoretical definition of Software development phases, it can be observed that there are no limitations when having to decide on a methodology for an RPA

specific project. Like any other project in the IT field, organizing the workflow and resources based on a specific methodology is the element that adds value to the implementation and may conclude in the success or failure of the project.

Software development methodologies, from traditional to agile approaches, are available for organizations to choose from. Even if some are taking a hybrid approach, their methodology preference is related to specific organizational projects and team traits. However, different sources cited in this paper are recommending implementing RPA in an Agile approach, as Agile principles are closely matching RPA criteria: the need to define functionalities from the beginning of the project, flexibility that allows improvements in the flow, opening to the stakeholders' feedback, being just a few.

By consolidating and analyzing knowledge from the research field, this paper provides valuable insight for organizations that want to achieve good RPA project management and improve their efficiency at the project level. The obtained results provide a foundation for continuing research in the direction of the most appropriate methodologies for RPA projects.

Despite the fact that the bibliometric analysis reveals that RPA project management research has experienced significant growth in recent years, indicating the rising interest in this field, based on the data acquired during this study project, it was determined that there is still insufficient information on this subject, at least in what scientific papers are concerned. However, as any other IT field, the creation and implementation of Robotic Process Automation initiatives depend heavily on the project lifecycles. Project lifecycles offer a methodical method of controlling RPA advances, ensuring that automation initiatives are carefully planned, carried out, and maintained throughout time. Even if the preference for the addressed methodology may not be directed towards Agile, the recommendations suggested as part of this study can be considered regardless, as these can be viewed as best practices for any RPA implementation.

The choice of a project management technique for RPA projects should ultimately be made based on the specific requirements and characteristics of the project, as well as the preferences and experience of the project team and other stakeholders. However, the available options to choose from do not differ from the ones that can be applied to any other development technology. The suggestions that have been documented as part of this research are following the general project life cycle of a project, and are giving an overview on some specific actions that can maximize RPA automations' efficiency.

Following these recommendations during various project phases can help RPA projects succeed, leading to increased process effectiveness, cost savings, and improved customer experience. These insights can help practitioners and researchers better navigate the challenges of RPA project management, as RPA is a continuously developing technology.

A future improvement would be to expand the area of analyzed papers with the use of multiple databases and additional sources, as the study was based only on the articles provided by the SCOPUS database (the recognized and highly valued source of scientific literature) and only included journal articles and conference proceedings.

References

- [1] Madakam, S., Holmukhe, R. M. and Jaiswal, D. K. (2019), "The future digital workforce: robotic process automation (RPA)", *JISTEM- Journal of Information Systems and Technology Management*, 16.
- [2] Gartner (2020), "Sustainable Business Strategy for a Positive Social and Environmental Impact", available at: <https://www.gartner.com/en/insights/sustainable-business> (accessed 16 December 2023).
- [3] Grand View Research (2022), "Robotic Process Automation Market Size, Share & Trends Analysis Report by Type, By Application, By Deployment, By Organization, By Region, And Segment Forecasts, 2023 – 2030", available at: <https://www.grandviewresearch.com/industry-analysis/robotic-process-automation-rpa-market> (accessed 23 April 2023).
- [4] GlobeNewswire (2022), "Robotic Process Automation Market Size Growing at 35.9% CAGR Set to Reach USD 25.1 Billion by 2030", available at: <https://www.globenewswire.com/news-release/2023/02/14/2608219/0/en/Robotic-Process-Automation-Market-Size-Growing-at-35-9-CAGR-Set-to-Rich-USD-25-1-Billion-By-2030.html> (accessed 23 April 2023).
- [5] Van der Aalst, W. M., Bichler, M. and Heinzl, A. (2018), "Robotic process automation", *Business & information systems engineering*, 60, 269-272.
- [6] Patrício, L., Ávila, P., Varela, L., Cruz-Cunha, M. M., Ferreira, L. P., Bastos, J., ... and Silva, J. (2023), "Literature review of decision models for the sustainable implementation of Robotic Process Automation", *Procedia Computer Science*, 219, 870-878.
- [7] Ostidick, N. (2016), "The Evolution of Robotic Process Automation (RPA): Past, Present, and Future", available at: <https://www.uipath.com/blog/rpa/the-evolution-of-rpa-past-present-and-future> (accessed 23 April 2023).
- [8] Gartner (2023), "Robotic Process Automation Software, Gartner Glossary", available at: <https://www.gartner.com/en/information-technology/glossary/robotic-process-automation-software> (accessed 8 April 2023).
- [9] Leshob, A., Bourgouin, A. and Renard, L. (2018, October), "Towards a process analysis approach to adopt robotic process automation", In 2018 IEEE 15th International Conference on E-Business engineering (ICEBE), IEEE, pp. 46-53.

- [10] Moreira, S., Mamede, H. S. and Santos, A. (2023), "Process automation using RPA—a literature review", *Procedia Computer Science*, 219, 244-254.
- [11] Lewin, C.(eds.) *Research methods in the social sciences*, London & Thousand Oaks CA: Sage Publications, pp. 41-49.
- [12] Sobczak, A. (2021). *Robotic Process Automation implementation, deployment approaches and success factors—an empirical study*. *Entrepreneurship and Sustainability Issues*, 8(4), 122.
- [13] El Khatib, M., Almarri, A., Almemari, A., & Alqassimi, A. (2023), "How Does Robotics Process Automation (RPA) Affect Project Management Practices", in *Advances in Internet of Things*, vol. 13 no. 2, pp. 13-30.
- [14] Blueprint (2020), "How an Agile approach to automation can deliver RPA at scale".
- [15] Ingo, P. (2021), "Agile RPA", available at: <https://scrumorg-website-prod.s3.amazonaws.com/drupal/2021-05/Agile%20RPA.pdf> (accessed 8 April 2023).
- [16] Mixson, E., (2022), "Why an agile approach is essential to scaling RPA", available at: <https://www.intelligentautomation.network/scaleable/articles/why-an-agile-approach-is-essential-to-scaling-rpa> (accessed 8 April 2023).
- [17] Hazzan, O. and Tomayko, J. (2004), "Human aspects of software engineering", *Extreme Programming and Agile Processes in Software Engineering: 5th International Conference, XP 2004*, Garmisch-Partenkirchen, Germany, June 6-10, 2004, Springer Berlin Heidelberg, *Proceedings 5*, pp. 303-311.
- [18] Fowler, M., & Highsmith, J. (2001). *The agile manifesto*. *Software development*, 9(8), 28-35.
- [19] Schwaber, K. and Sutherland, J. (2020), "The Scrum Guide", *Scrum Alliance*, 21(1), 1-38.
- [20] Schwaber, K. (2004), *Agile Project Management with Scrum*, Microsoft Press.
- [21] Atlassian. *Agile Teams*. Retrieved from: <https://www.atlassian.com/agile/teams>.
- [22] Murch, R. (2012), *The Software Development Lifecycle-A Complete Guide*, Kindle Edition.
- [23] Cha, J. and Maytorena-Sanchez, E. (2019), "Prioritizing project management competences across the software project life cycle", in *International Journal of Managing Projects in Business*, vol 12, no. 4, pp. 961-978, doi:10.1108/ijmpb-11-2017-0145.
- [24] Haskins, B., Stecklein, J., Dick, B., Moroney, G., Lovell, R. and Dabney, J. (2004), "8.4.2 Error Cost Escalation Through the Project Life Cycle", in *INCOSE International Symposium*, vol. 14 no. 1, pp. 1723-1737.
- [25] Sobczak A. (2022), *Robotic Process Automation as a Digital Transformation Tool for Increasing Organizational Resilience in Polish Enterprises*, *Sustainability*,14(3):1333. <https://doi.org/10.3390/su14031333>.
- [26] Willcocks, L. P., Lacity, M. and Craig, A. (2015), "The IT function and robotic process automation", *The Outsourcing Unit Working Research Paper Series*, Paper 15/05.
- [27] Bygstad, B. (2015), "The Coming of Lightweight IT", *ECIS 2015 Completed Research Papers*, Paper 22, ISBN 978-3-00-050284-2.
- [28] Rosen, M., Lublinsky, B., Smith, K. T. and Balcer, M. J. (2008), *Applied SOA: Service-Oriented Architecture and Design Strategies*, Indianapolis, John Wiley and Sons.
- [29] Moayed, V. (2017), "From pilot to full scale RPA deployment", *UIPath whitepapers*, available at: <https://cdn2.hubspot.net/hubfs/416323/>

- Whitepapers/From%20pilot%20to%20full-scale%20RPA.pdf?t=1509016640696 (accessed 10 May 2023).
- [30] Deloitte (2018), “Understanding the challenge of implementing your virtual workforce”, available at: <https://www2.deloitte.com/content/dam/Deloitte/nl/Documents/strategy/deloitte-nl-so-understanding-challenge-of-implementing-rpa.pdf> (accessed 12 April 2023).
- [31] Kristel, T., & Miller Pamela, F. (2018). „Research Management “.
- [32] VOSViewer, available at <https://www.vosviewer.com/> (accessed 10 May 2023).
- [33] Bostan, A-I and Virgolici, O. (2023), “A study into the implementation of Robotic Process Automation in different business processes”, in Proceedings of the 17th International Conference on Business Excellence, Rethinking business: Sustainable leadership in a VUCA world, 23-25 March 2023, Bucharest, Romania.
- [34] Barbour, R. S. and Schostak, J. (2005), "Interviewing and focus groups", in Research methods in the social sciences, 1(1), 41-48.
- [35] Project Management Institute (2018), “Pulse of the Profession”, available at: <https://www.pmi.org/-/media/pmi/documents/public/pdf/learning/thought-leadership/pulse/pulse-of-the-profession-2018.pdf> (accessed 10 May 2023).



Andreea Izabela BOSTAN is a PhD student at the Academy of Economic Studies from Bucharest, her research topic being “Business processes optimization through automation, in the context of software project management”. She graduated the Master’s degree program Information System for Economic Process and Resources Management (SIMPRES) in 2021. Her fields of interest are agile methodologies, RPA and project management.



Ana Ramona BOLOGA graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 1999. She got the PhD in Economics in 2007. At present she is a professor in the department of the Faculty of Cybernetics, Statistics and Economic Informatics and Cybernetics Department of the Academy of Economic Studies of Bucharest. Her fields of interest are: integrated information systems, information system analysis and design methodologies, and software agents.