POLYTECHNIC UNIVERSITY OF TURIN

Master Degree Thesis

Artificial intelligence for generation and verification of BPMN diagrams

Supervisors
prof. Riccardo Coppola
doct. Giacomo Garaccione

Candidates
Anxhela Asllani
matricola: s288451

Academic Year 2023-2024

Abstract

Software modelling has grown in value since the end of the 20th century, and its importance continues to develop with IT systems becoming increasingly prominent, complex, and larger in scale.

This dissertation addresses the intersection between artificial intelligence (AI) and software modelling, with a focus on the generation of Business Process Model and Notation (BPMN) diagrams by AI. Being that BPMN helps effectively communicate business processes, this research is focused on the potential capability of AI, specifically GPT 3.5 and GPT 4.0, to generate and create these diagrams automatically starting from descriptive problem statements.

By creating a scoring system to evaluate diagrams and syntax errors in the solutions generated by the three actors (human, GPT 3.5 and GPT 4.0) as well as using statistical analyses such as t-tests and ANOVA, the research aims to assess whether AI-generated solutions can achieve or even exceed the efficiency of human-generated solutions in terms of accuracy and speed. The final statistical results suggest that GPT 3.5 and GPT 4.0 do not outperform human performance in generating BPMN diagrams.

This outcome highlights the difficulties and limitations of GPT(Generative Pre-trained models) and Large Language Models in apprehending and deploying complex contexts and requirements, generating graphic representations, and possessing limited domain-specific knowledge. It also draws attention to the importance of ongoing exploration and technological progress to enhance AI capabilities in this field.

Acknowledgements

I would like to start this dissertation by extending my special thanks to all those who helped me complete this work and this important journey. Firstly, I would like to express heartfelt thanks to Prof. Riccardo Coppola, who followed and supervised the process of this work, and to my co-mentor, Doct. Giacomo Garaccione, who was constantly ready and available to help and advise through each stage of the work. Their professional experience and academic advice were invaluable in giving the right direction to this paper.

Secondly, I would like to thank my manager and work colleagues for their understanding and support. Reconciling work and study was not easy, but they definitely made it easier for me.

I would like to express my gratitude and thank all my close friends for the inspiration and motivation they gave me during good times and hardships.

Last but foremost, I want to thank my parents and family for their unconditional love, trust, and dedication throughout this research and all these years.

Contents

List of Tables 5

List of Figures 6

1 Introduction 9

2 Background 11

2.1 Definition of Software Modelling 11
2.2 Modelling with UML Class Diagram 13
2.3 Definition of Process Modelling 20
2.3.1 Definition, Characteristics and Notions of Business Process Modelling 21
2.3.2 Formats available For BPMN Diagrams 25
2.4 Large Language Models and Conversational Artificial Intelligence 28
2.4.1 Introduction to Large Language Models 28
2.4.2 Conversational Artificial Intelligence and Introduction to OpenAI and ChatGPT 29
2.4.3 Advantages and disadvantages related to the technology 31
2.4.4 Ethical Concerns Related to AI 32

3 AI Applications 35

3.1 AI applications in Education and Academia 35
3.2 AI applications in Industry 36
3.3 AI applications in Finance 38
3.4 AI applications in Medicine and Healthcare 40

4 Methodology 43

4.1 Introduction 43
4.2 Background and Literature Review 43

4.3 Study Design 45
4.4 Prompt Design 47
4.4.1 Prompt Development Process for BPMN Diagrams . . 48
4.5 Evaluation Methodology 52
4.5.1 Evaluation Criteria 52
4.5.2 Evaluation Process 53
4.5.3 Sample Evaluation 57
4.5.4 Statistical Methods 60
4.5.5 Limitations of the Evaluation Methodology 62

5 Results and Discussion 65
5.1 Results 65
5.2 Discussion 69

6 Evaluations 73
6.1 Evaluation of AI-generated diagrams 73
6.2 Comparative Evaluation 76

7 Conclusions and Future Work 79
7.1 Conclusions 79
7.2 Future Work 80

Bibliography 81

List of Tables

2.1 Advantages and disadvantages of AI 31
4.1 Scoring System 54
4.2 Syntax Errors 56
4.3 Reporting format for syntax errors found in the diagrams . . . 57
6.1 Most common logical errors found in the GPT-made diagrams 76

List of Figures

2.1 UML Class Representation 14
2.2 UML Class Diagram Relationships.Source:[2] 15
2.3 Notions of business process modelling: 21
2.4 Types of tasks in BPMN 23
2.5 Types of gateways in BPMN 23
2.6 Types of events in BPMN 24
2.7 Sequence flow arrow 24
2.8 Pools and Lanes in BPMN 25
2.9 Ethical Concerns of AI technology 33
4.1 Evaluation Process 47
4.2 GPEI Methodology, Source:[52] 49
4.3 Initial Prompt 49
4.4 Textual Representation of solutions generated by GPT 3.5 . . 50
4.5 Prompt Refinement 51
4.6 Final Prompt 51
4.7 Matrix of Evaluation Criteria 55
4.8 Final Matrix 56
4.9 Solution Vs. Human 58
4.10 Evaluation Scores Human vs. Solutions 58
4.11 GPT 3.5 GPT 4.0 vs Solution 59
4.12 Final Scores 59
4.13 Structure of the appropriateness evaluation dataset for each exercise 60
4.14 Final evaluation dataset 61
5.1 t-test Human vs GPT 3.5 66
5.2 t-test Human vs GPT 3.5 67
5.3 Anova Analysis Accuracy scores 68
5.4 ANOVA analysis syntax errors. 68
6.1 Diagram generated from model GPT 3.5 74
6.2 Diagram generated from model GPT 4.0 75

6.3 Accuracy percentages between actors 77
6.4 Syntax errors percentages 77

Chapter 1

Introduction

Since the middle of the 20th century, the field of software engineering has undergone a rapid transformation. In this context, software modelling has become an important pillar of IT systems since they are becoming larger, more complex and more comprehensive. Two important standards in the field of software modelling are UML diagrams and BPMN diagrams, which help reflect the components of information systems and the constituent processes of an organization.

Usually, these diagrams are developed in the study phase of the IT system that must be implemented and are realized through a repeated process as they involve several experts and are time-consuming. Regardless of this, the possibility that these diagrams contain errors is not negligible. In this context, the question of whether the capacity and potential of artificial intelligence models can be used to make this process faster and more accurate becomes important. Automation of this process would be more important in the IT industry.

The goal of this study is to test whether the GPT 3.5 and GPT 4.0 models can generate BPMN diagrams starting from the initial description of the problem. After generating the diagrams, an empirical comparison between the human-generated diagrams and those generated by the two AI models is needed. This process is challenging since, at this moment, there is little evidence of whether the specific knowledge of AI models in the BPMN domain and its capability to create graphical representations are sufficient to give reliable results.

This thesis includes a comprehensive approach combining existing research works in the field of process modeling and providing empirical results experimental work. By analyzing a problem description set and directing the model to generate diagrams once from the 3.5 model and once from the 4.0 model, the results should be compared to the human-generated diagram. The evaluation criteria aim to elaborate and compare the aspect of accuracy and the aspect of syntactic errors. Starting from the null hypothesis that there is no difference in the performance of the three actors in the generation of diagrams, statistical analyzes will have to provide a scientific proof of the truth of this statement.

In summary, this study has several main objectives:

Creating a rating system to evaluate human and AI-generated diagrams, focusing on accuracy and syntax compliance. Using statistical tools such as T-tests and ANOVA to test the null hypothesis Exploring current AI capacity in analyzing and generating process modelling diagrams via XML files.

The structure of this paper will be as follows: The second chapter will give a background and the main notions of software modelling and those of large language models; the third chapter will give an overview of the latest AI developments in different sectors; the fourth chapter will focus on the methodology, the fifth chapter will reflect the results achieved and their interpretation; the sixth chapter will give a summary of the performance of the AI and human in a more synthetic way, and in conclusion, the last chapter will summarize the final conclusions of this paper.

The aim of this paper is to enrich the conversation on the role of AI in the field of process and software modelling. The accumulated conclusions will not only provide an answer to where GPT models currently stand in this domain, but will also pave the way for future research to unlock the full potential of AI in software modeling.

Chapter 2

Background

2.1 Definition of Software Modelling

The Oxford Dictionary defines modelling as the work of creating a simple description of a system/process that can be used to explain it and predict outcomes, or simply as the activity of creating models of objects [1]. Since the dawn of humanity, ancient civilizations have used modelling as an important element to represent an architectural idea and as a verification tool to find the best possible idea. As a concept, modelling is used extensively in the software field, so-called software modelling, with the aim of simplifying and making complex software systems less abstract. Software engineering took off in the late 1960s and continues to develop today. The change in approach is due to the fact that previously the focus of software engineering was merely on code generation, without any conventions on processes or techniques. Becoming an increasingly important part of all the fields around us, computer systems have become more complex and comprehensive. So in light of the arguments aforementioned, modelling techniques have become a critical part of better understanding software systems.

The main purpose of software modelling is to understand, design and document different integral aspects of the system, thus having a key role in all phases of the software development life cycle, from feasibility studies, functionality, trade-offs, features, design and cost analysis including every other aspect of the components.

According to the Object Modeling Group (OMG), "Modelling is the design of software applications before coding."[2] As mentioned above, modelling is not just about code generation, but about every stage of the system development cycle, starting from requirements gathering, requirements analysis, system design, implementation and testing the final results. Another form of software modelling has to do with its abstract nature. As technologies are continuously developing, systems evolve dynamically, making it necessary for the model itself not only to contain one logic or algorithm but to also contain several specifications presented such as models, diagrams or notations. There are different types of software models, thus distinguishing structural, behavioural and functional models, all with different purposes to describe and present different specifications of the system they represent.

These models, in their specifics, are used as an instrument to improve the communication between stakeholders, developers, product designers and customers. Furthermore, they help to better understand the functioning of the system, how it can be created, how it can be developed, how it can be improved and what are the technologies that compose it and the communication with external systems. The role of software modelling starts from problem identification, solution exploration, and requirements analysis. In this way are explored different facets of the system throughout the software development life cycle, thus ensuring maximum reliability and quality of software development. Furthermore, the key part of the process is the documentation of the system design and architecture, thus providing a reference for continuous developments or improvements. This ensures transparency and simplifies communication between stakeholders.

Software models can be built using different techniques depending on the functionality to be presented, mentioned here: flowcharts, state transition diagrams, UML diagrams and SysML diagrams. UML is a language used to build diagrams, while SysML is an extension of UML: both are part of software modelling and are used to describe different parts of the system. The criteria for choosing appropriate modelling tools take into account three main criteria. The first criterion has to do with the requirements of the project, starting from the purpose and objective that you intend to achieve with this project, the volume and flow of data and the integration of this tool with software or other platforms that make part of the project in question. The second criterion takes into account scalability and how the tool is able to adapt to the continuous increase in data. The third criterion considers the complexity of the tool, its costs and its flexibility.

Overall, software modelling is a vital part of the software development lifecycle. It not only helps in the quality and reliability of the system but plays an important role in streamlining communication between stakeholders, developers, business analysts and functional analysts thus ensuring effective communication between them. To accomplish this, conventional methods, techniques, notations, and diagrams play a critical role in representing the presentation of different forms of the system, such as UML and SysML[3].

2.2 Modelling with UML Class Diagram

The Unified Modeling Language (acronym as UML) is a standardised visual modelling language that is mostly used in software engineering and it is an object-oriented language. It is used to design, visualise and document software systems throughout its lifecycle. In the early 1990s, in the field of software engineering, there were various modelling languages existing in this industry, that is the reason why emerged the need for a unified modelling language with a unified syntax and semantics. The first version of UML (UML 1.0), was released in 1997 and it answered the need to provide a standardised notation and semantics to visualise, specify, build and document software systems. In the later years, UML refined and provided enhancements based on industry and advancements in the software industry[4].

The diagrams that are included in UML are the following:

1. Use case diagrams
2. Class diagrams
3. Object diagrams
4. Communication diagrams
5. Sequence diagrams
6. State Machine diagrams
7. Activity diagrams

Among all these different diagram types, class diagrams are the most relevant. Class diagrams help create a visual representation of the parts and actors included in the organisation in the system. They help as well to show the connections within that system or application by portraying every element as an object. Before explaining the way in which a class diagram works, it is important to mention two key concepts: *Classes* (the type) and *Objects* (an instance of the type). These notions are depicted as boxes in the UML notation, as shown in Figure 2.1

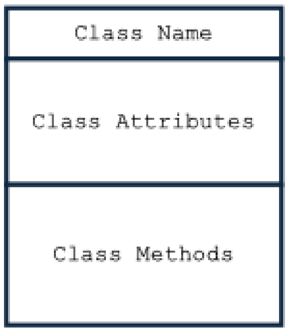

Figure 2.1. UML Class Representation

The above picture gives an illustrative example of the class where the top part of the rectangle box represents the class name, the middle part holds the attributes (variables or properties) of the class, while the bottom part holds the operations, called methods or functions. A sophisticated software system is depicted as a set of classes that are connected between them by lines in which their static relationships are displayed. There are various types of relationships used to connect these classes, and each of them expresses a different type of relationship and has a different role. The most frequently used are:

1. Associations
2. Aggregation/Composition Hierarchy
3. Generalisation/Specialisation

An association is a structural relationship between two or more classes, represented by a line that joins the classes that need to be connected, with the possibility of having a small arrow at the end of the line, to show the direction in which the association between these two classes should be perceived.

Another element of associations is their multiplicity, a number used to show how many instances of a class can be connected to instances of another class: a *1* means that an instance of a class is connected to exactly one instance of another class, *0..1* means an optional connection with exactly one instance of a second class, *1..** means that the class must be connected with at least one instance and may be connected with more of them, and so on (different types of multiplicities are shown at the top right of Figure 2.2).

Aggregation and composition hierarchies are whole/part relationships. This form of connection is represented by a diamond icon with differences in how they are used. Aggregation is represented by a blank diamond icon and means that if the parent class is deleted, the child class may exist. In contrast, the composition is reflected by a filled diamond and describes the case where the child class cannot exist if the parent class is deleted from the system.

Another type of relationship is the generalisation/specialisation hierarchy. This connection is hereditary, meaning that the child class inherits all the specifications of the parent class. It is represented by a line icon with a closed arrow pointing from the child class to the parent class[3].

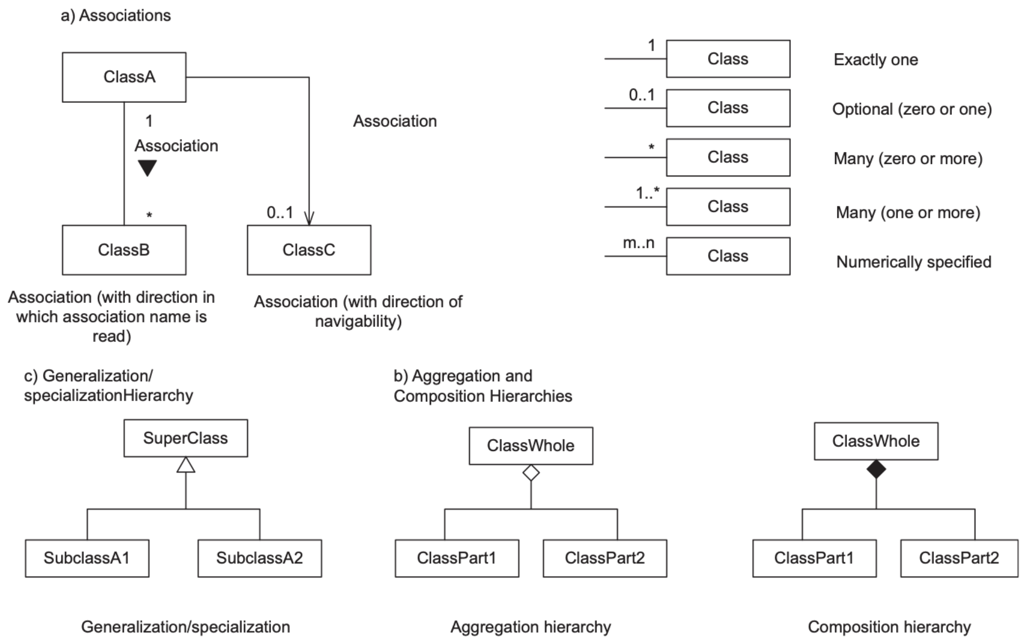

Figure 2.2. UML Class Diagram Relationships.Source:[2]

Another important characteristic of UML class diagrams is that they are static models, meaning that the main components of the diagrams do not vary over time. Unfortunately, this means that the dynamic information flow and the interactions between classes cannot be interpreted from a UML class diagram.

To sum up,the logical steps that are used when building UML class diagrams are as follows:

1. Class Identification
2. Definition of Attributes

3. Definition of Methods/Functions
4. Establishment of relationships
5. Perfectioning of the Model
6. Documentation
7. Consolidation with other UML diagrams

Firstly, it is important to identify what are the main classes that will compose the diagram. The diagrams are usually built by experts, and it is a process where several ideas are brainstormed. The first details defined are classes, their characterizing attributes, and the methods that they perform; after that, relationships with other classes are defined.

Diagrams can then be perfected over time according to business needs, starting from a low-level perspective until an encompassing high-level view of the system is created[3].

Formats available to model with UML class Diagrams

This section will discuss how UML class diagrams can be exchanged in many other different system documents. In recent years, the so-called UXF (UML exchange format) has emerged as a popular strategy to exchange UML class diagrams in different system documents[5]. Examples of exchange formats are:

1. Image File Format
2. Vector Graphics Format
3. UML Files
4. XML Format

The choice of format is influenced by many key decisions, beginning from the diagram's intended application, its compatibility with other complementary systems/tools that need to be used based on the case, the tool's capabilities, and personal preferences. It is very important to take into account additional tools or systems that might need to interact with the UML class diagram when choosing a format[5].

The focus of this section is to describe and interpret how a UML class diagram can be converted into an XML Schema, the current standard for

Internet documents. The aim of transforming the UML class diagram into an XML schema is to have the diagram in an easily readable, transformable, and reusable format. The conversion to an XML schema can give a flexible method to represent the logical structure of the document in various ways.

Depending on the individual standard or schema, an XML structure or schema that represents a general case of a UML class diagram can change based on the complexity of the specificity of the diagram itself. The following example illustrates a general and simple example of how an XML structure may be built:

```
<uml:ClassDiagram xmlns:uml="http://www.omg.org/spec/UML
    /20131001" xmlns:xmi="http://www.omg.org/spec/XMI/20131001">

  <packagedElement xmi:type="uml:Package" name="Package_Name">

    <!-- Class Definitions -->

    <packagedElement xmi:type="uml:Class" name="Class_Name">

      <!-- Class Properties -->

      <ownedAttribute xmi:type="uml:Property" name="Attribute_Name
          " type="DataType" visibility="visibility_type"/>

      <!-- Class Operations -->

      <ownedOperation xmi:type="uml:Operation" name="
          Operation_Name" visibility="visibility_type">

        <!-- Operation Parameters -->

        <ownedParameter xmi:type="uml:Parameter" name="
            Parameter_Name" type="Parameter_Type"/>

      </ownedOperation>

    </packagedElement>

    <!-- Associations -->
```

```
    <packagedElement xmi:type="uml:Association" name="
        Association_Name">

      <!-- Association Endpoints -->

      <memberEnd xmi:type="uml:Property" type="Class1"/>
      <memberEnd xmi:type="uml:Property" type="Class2"/>

    </packagedElement>

    <!-- Generalizations/Inheritances -->

    <packagedElement xmi:type="uml:Generalization" general="
        Superclass" specific="Subclass"/>

    <!-- Aggregations -->

    <packagedElement xmi:type="uml:Association" name="
        Aggregation_Name">

      <!-- Aggregation Endpoints -->

      <memberEnd xmi:type="uml:Property" type="Class1" aggregation=
          "shared"/>
      <memberEnd xmi:type="uml:Property" type="Class2" aggregation=
          "composite"/>

    </packagedElement>

    <!-- Dependencies -->

    <packagedElement xmi:type="uml:Dependency" client="Class1"
        supplier="Class2"/>
    <!-- Other Elements: Interfaces, Packages, etc. -->

  </packagedElement>

</uml:ClassDiagram>
```

The XML structure complies with the UML metamodel. As it can be distinguished in the structure shown above, in the xml structure there are components like: *uml:Class*, *uml:Property*, *uml:Operation*, *uml:Association*, *uml:Generalization*, *uml:Dependency*, with each of them representing a distinct element in the diagram.

The first part of the schema is composed of the namespaces: the *xmlns:uml*' and *xmlns:xmi* attributes respectively define the namespaces for the UML and XMI standards.

Moreover, the *<packagedElement>* element in the schema represents a package which acts as a container for the diagram's classes, associations, generalisations, dependencies and all its essential elements.

According to this definition, the *<packagedElement xmi:type="uml:Class" name="Class_Name"* node represents a class in the UML class diagram: the node includes components that define the properties or attributes of the class, as well as its operations.

After having defined the classes, the attributes and operations of each of them need to be specified. The *<ownedAttribute>* element represents a class property (attribute) by specifying the name, type, visibility, and other properties of an attribute associated with a class. The *<ownedOperation>* element represents a class operation (method) and defines the name, visibility, and parameters of an operation associated with a class; operation parameters are represented using the *<ownedParameter>* element.

Other component parts when building a UML class diagram are the associations, generalisation and aggregation between classes, and in an XML schema they are represented as follows:

- A *<packagedElement>* element with *xmi:type="uml:Association"* represents an association between classes in the UML class diagram; it typically consists of two or more *<memberEnd>* elements, each specifying the participating classes and the role they play in the association.

- A *<packagedElement>* element with *xmi:type="uml:Generalization"* represents a generalisation or inheritance relationship between classes and specifies the parent class ('general') and the child class ('specific') involved in the inheritance.

- Aggregation relationships, such as composition and shared aggregation, are represented using *<packagedElement>* elements with *xmi:type="uml:Association"* and *<memberEnd>* elements with an *aggregation* attribute that specifies the type of aggregation.

It should be noted that the XML structure provided is a general representation and may not include all elements. For describing UML class diagrams in XML, different XML schemas, tools, or UML versions may have additional elements or particular needs based on the specific schema to be represented and its scale of complexity.

2.3 Definition of Process Modelling

A process is described as a chain of actions that, by giving a specific input, produce a specific output, and that could be used to represent a system at various levels of detail. Before trying to represent a process, it is essential to know the activities included and the order in which they are executed. The other side of the coin of this argument is the fact that, in some cases, some activities are really difficult to distinguish from others and this happens mostly when they are strictly related to one another.

In the field of business importance, process modelling is defined as the development of a visual schema in order to represent or describe a system or a process. This activity involves the documentation of all the actions, steps and analyses that are conducted during this phase in order to comprehend and evaluate them better. The most used areas are project management, software development, engineering, and business process management.

The main goal of process modelling is to communicate to stakeholders a clear and organised view of the way in which a process operates. Hence, in this manner, they can evaluate how the process works, suggest further enhancements and optimise it.

Process models and diagrams can be used to design and plan future states/improvements (to-be) as well as to communicate and document the current state of a process (as-is).

Some examples of process models diagrams are flowcharts, business process models, and data flow diagrams. These models depict various components and activities that compose a process such as tasks, decisions points, inputs, outputs, and dependencies by using a variety of symbols, notations, and diagrams all specific to the type of formalism being used.

2.3.1 Definition, Characteristics and Notions of Business Process Modelling

Processes are a substantial part of business whether the latter provides a service or a product to the client. The design and execution of processes impact in a direct way the quality of service perceived from customers and its effectiveness. As a matter of fact, if one organisation has better processes and executes them better, it can surpass another competitor that provides the same/similar types of services.

Business process modelling (BPM) is a formalism that represents business processes in a visual form and allows for analysing the activities, interactions, and information flows within an organisation's business processes.

Firstly, it involves creating graphical representations of processes to understand their functioning better: this helps identify the improvement areas, and enhance communication and collaboration among interested stakeholders.

Secondly, another critical aspect of business process modelling is its capability to represent a process using diagrams and symbols. This makes it easier to understand and analyse processes more efficiently, no matter how complex they are.

Thirdly, BPM has a process-oriented nature, meaning that its duty is to capture the sequence of activities and their relationships within a process. This includes understanding information flows and tasks involved in the process by keeping the focus on the high-level representation of the process and not giving information on details that regard to the implementation.

As the process becomes more convoluted, the diagram gets refined iteratively by insights and suggestions provided by stakeholders, ensuring that the model remains coherent and updated and that it communicates correctly the representation of the process[6, 7].

Figure 2.3 displays the most essential notions of business process modelling in their intended order:

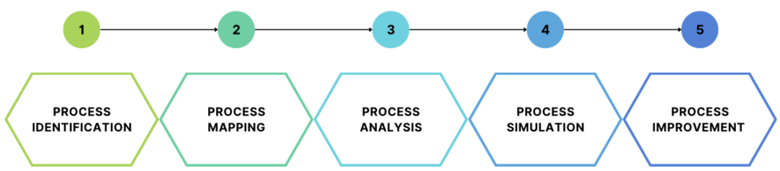

Figure 2.3. Notions of business process modelling:

The first step in modelling a process is identifying the organisation's indispensable processes; understanding the process's limitations and the type of activities involved is critical at this step.

In the second phase, business analysts create visual representations (such as flow charts or diagrams): these schemas should portray the order of events, decision points, and information flow inside the process being analysed, allowing for the detection of bottlenecks, inefficiencies, and enhancements.

In the third step, the stakeholders use several procedures to find areas that can be optimised, with the final objective being an increase in the process' total effectiveness.

Following this phase of the analysis, the field experts use simulation techniques to reproduce the process's behaviour. After that, stakeholders understand the process's performance by analysing different scenarios and identifying potential inefficiencies or opportunities for improvement.

The goal of process improvement is to spot and address solutions to the faults or inconsistencies found in the fourth step. Experts propose several changes that can include redesigning the process, for example usage of new tools or technologies, or best case scenarios from other analogue cases.

BPMN was created from the lack of a single standard for communicating the processes that make up a company. In the early 2000s, there were different notations and points of view regarding the representation of company processes, and different companies used different methods. Finding themselves in this situation, industry experts thought of creating a common standard notation to create a single standard and then to facilitate the communication of business processes between different stakeholders. In May 2004, the BPMN 1.0 notation was released, and in the following years, this standard was improved until 2011, when the BPMN 2.0 standard was published. This international standard now succinctly embodies best practices in business modelling [8].

BPMN (Business Process Model and Notation) is a formalism that serves as a standard for modelling business processes. Its primary purpose is to assist technical and business users in business process modelling. Precisely for this reason, the BPMN standard uses simple, intuitive, and understandable notations. Based on simple flowchart techniques, BPMN can communicate complex systems clearly, facilitating communication between business analysts and technical developers [8].

Business process modelling language, like every other modelling language, has its own notations and symbols used to represent different artefacts on process within the business reality. It is important to give a clear description of the notations and symbols used in this formalism. The most used components are:

1. Tasks: they represent activities or actions carried out by actors involved in the business processes. Its shape is a rectangle box with inside the action that is being performed[9].

Figure 2.4. Types of tasks in BPMN

2. Gateways: they represent decision points and are represented with a diamond shaped symbol. The gateways are of different types: exclusive (only one outgoing flow can continue), event-based or parallel (all outgoing flows continue and rejoin at some point in the process), with each symbol representing a distinct logical condition[9].

Figure 2.5. Types of gateways in BPMN

3. Events: they represent something that can happen at a certain moment during the business process and they are reflected as a circle shape. Events can be of start type (thin circle), end type (thick circle) or intermediate type (double circle). These events sometimes might contain in the middle of the circle an icon that represents their activation condition (e.g. an error, a specific moment in time, an error instance, or sending/receiving a message)[9].

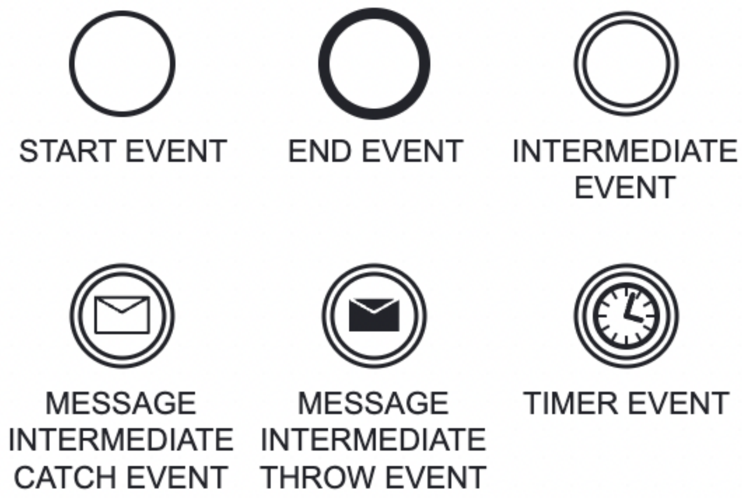

Figure 2.6. Types of events in BPMN

4. Flows: they are used to mirror the activity flows between different activities and/or participants and can be of two types: sequence and message. Sequence flows are represented as a continuous arrow and indicate the sequence in which the different activities happen by connecting the source and destination elements. On the other hand, message flows are used to represent the exchange of messages within the system and are represented as a dashed arrow. While sequence flows show the order of the activities of the same participant, message flows are used to show the order of activities of different participants and information exchanges among them[9].

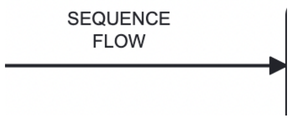

Figure 2.7. Sequence flow arrow

5. Pools and lanes: they are represented by a rectangle-shaped form and are used to identify participants/entities involved in the process. Pools differ from lanes because the first represents companies, departments or

even roles while lanes represent sub-roles within the macro entity that is defined by the pool name [9].

Figure 2.8. Pools and Lanes in BPMN

These are only some of the types of objects used to build a BPMN diagram. These diagrams are read from left to right, following the direction of the sequence flows; in case of message flows that connect different pools, the reading is done following their direction (either top-bottom or vice-versa).

2.3.2 Formats available For BPMN Diagrams

As previously discussed in the UML case, BPMN diagrams can be saved in different ways depending on the user's preferences and their necessities; since these schemas are usually created with specific tools or software, it depends mostly on the type of software being used and what it allows the user.

Examples of formats available for BPMN diagrams are:

1. BPMN XML Schema
2. Image Format
3. Interactive Web Formats
4. BPMN Diagramming Tools

It is the benefit of the study to follow the BPMN 2.0 standard when trying to create an XML file that depicts the BPMN diagram. The structure of the XML schema consists of two parts: the first one is the graphical information specific to the software/tool being used, and the second one is the structural information, that details the structural components of the diagram. The example below shows only the structural details and not the graphical part.

```
``` xml
<?xml version="1.0" encoding="UTF-8"?>
<definitions xmlns="http://www.omg.org/spec/BPMN/20100524/
    MODEL"
             xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
             xsi:schemaLocation="http://www.omg.org/spec/BPMN
                 /20100524/MODEL BPMN20.xsd"
             id="definitions_1"
             targetNamespace="http://bpmn.io/schema/bpmn"
             exporter="YourExporter"
             exporterVersion="1.0">

  <!-- Define the process -->

  <process id="process_1" isExecutable="true">

    <!-- Define the start event -->

    <startEvent id="startEvent_1">

      <!-- Additional properties or attributes for the start event -->

    </startEvent>

    <!-- Define other BPMN elements like tasks, gateways, and events
        -->

    <!-- Define the end event -->

    <endEvent id="endEvent_1">

      <!-- Additional properties or attributes for the end event -->

```
```

```
    </endEvent>

    <!-- Define sequence flows to connect the elements -->

    <sequenceFlow id="sequenceFlow_1" sourceRef="startEvent_1"
        targetRef="task_1" />

    <sequenceFlow id="sequenceFlow_2" sourceRef="task_1" targetRef="
        endEvent_1" />

    <!-- Define other sequence flows as needed -->

  </process>

</definitions>
```

The first part of the BPMN schema is the section called *<definitions>*, which defines the most important elements like BPMN namespace and schema.

The additional elements define the structural components of the diagram, such as processes (*<process>*), sequence flows (*<sequenceFlow>*), start events (*<startEvent>*), end events (*<endEvent>*) and the like. Each item should have a unique *<id>* and possibly the boolean attribute *<isExecutable>*.

The latter determines whether the presented process is executable or not by the BPMN tool. For example: In this case, (*<isExecutable>*) is set as true, which means that the process described in the XML schema can be executed by the BPMN software reading the file and it is intended for direct implementation. In the opposite case, when this attribute is set to False, it means that the process is not executable, but is only used for documentation and cannot be implemented.

Following the same logic, the other parts of the diagram are also defined in the same way from a syntax perspective, but based on their specificities, they might contain other attributes. For example, the element *<sequenceFlow>* defines the message flows, so besides the id attribute, it should also specify the source and destination elements.

Moreover, if gateways are present in the diagram, they are defined through the *<conditionExpression>* element within the *<sequenceFlow>*. In the same syntax, other elements of the diagram such as timer events (*<timerEventDefinition>*), message events (*<messageEventDefinition>*) or other parts might be defined as well, adding several case-specific attributes.

The XML schema illustrated above is only a simple demonstration of how it is built as an XML file, but its elements depend on the complexity of the diagram: the more complex the diagram, the richer the detail of the schema.

Finally, the end part of the schema contains the graphical definitions and the coordinates of each part of the process, but this changes based on the software or tool used.

2.4 Large Language Models and Conversational Artificial Intelligence

2.4.1 Introduction to Large Language Models

Large Language models (LLM) can be considered as an instance of a foundation model. Foundation Models are pre-trained on large amounts of unlabeled and self-supervised data, meaning that the model learns from patterns in the data in a way that produces generative and adaptable output. Thus, an LLM is an instance of FM applied specifically to text and text-like things (i.e. code).

An important remark to be made is that LLMs are trained on large datasets of text, such as books, articles and conversations that might reach up to 1 PB of data processed.

Large Language Models are Large, Generative purpose language models that can be pre-trained and then fine-tuned for specific purposes. LLMs are used to solve common language problems such as text classification, question answering, document summarization and text generation. Moreover, these models can be trained to solve different problems in retail, finance and entertainment areas.

By the term "Large" it is intended the fact that the model has an enormous size of training data set (PetaByte Size) and that it has a large number of parameters. A parameter is a value that the model can change independently as it learns and the more parameters a model has the more complex it can be[10].

In machine learning, parameters are often called hyperparameters: they are the memory and the knowledge that the machine learned from model training and they define the skill of a model in solving a problem such as predictive text.

Another characteristic of LLMs is their transformative architecture that enables the model to handle data sequences like sentences or lines of code. Transformers are designed to understand the context of each word in a sentence by considering it in relation to every other word. This allows the model to build a comprehensive understanding of the sentence and the meaning of the words within it[11].

Subsequently, the architecture is trained on this large amount of data. During training, the model learns to predict the next word in a sentence, but with each iteration, the model adjusts its internal parameters to reduce the difference between its predictions and the actual outcomes. The model keeps doing this gradually, improving its word predictions until it can reliably generate coherent sentences,

Furthermore, the model can be fine-tuned in a smaller and more specific dataset and by doing this it refines its understanding to be able to perform a specific task more accurately. Fine-tuning is what allows a general language model to become an expert at a specific task.

There are 3 main types of Large Language Models:

1. Generic (Raw) LM: These models predict the next word (technically token) based on the language in the training data.

2. Instruction Tuned LM: The model is trained to predict a response to the instructions given in the input.

3. Dialog Tuned LM: The model is trained to have a dialogue by predicting the next response.

Each of these language models needs prompting in a different way[12].

2.4.2 Conversational Artificial Intelligence and Introduction to OpenAI and ChatGPT

Conversational artificial intelligence (AI) is a rapidly developing technology that has had a significant impact on e-commerce, education, entertainment, health, productivity, and journalism. As a result, businesses, governments, and academic institutions are becoming more and more interested in this domain[13].

It uses a large amount of data (as well as a large volume of machine learning) with the aim to help imitate human interactions. Moreover, it can recognize speech and text inputs and translate their meanings across various languages meaning that it generates a response based on user intent. It is important to emphasise the fact that intents are different from entities: intents allow a machine to to decipher what the user is asking, while entities are used to act as a way to provide relevant responses.

Conversational Artificial Intelligence (Conversational AI) is a branch of Artificial Intelligence that focuses on creating AI agents capable of replicating and automating human conversations and linguistic interactions. It makes use of developments in natural language processing (NLP), to enable machines to comprehend and react to human speech and text. Various industries, including aviation, tourism, and healthcare, have implemented conversational AI technologies, such as chatbots and virtual assistants, to improve customer management and improve user experience.

There is yet little use of conversational AI in the architecture, engineering, and construction (AEC) sector, and little research has been done in this field. However, by boosting productivity, enhancing user experience, and assisting numerous jobs across the project lifecycle, conversational AI holds enormous promise for revolutionising the AEC business[14].

There are four basic steps to the natural language processing that occurs within a conversational AI:

1. Input Generation: Users provide input in the form of voice or/and text through a website or an app.

2. Input Analysis: Used to decipher the meaning and derive its intent through Natural language understanding

3. Dialog Management: Used to formulate a response in a way that mimics human speech using a Natural Language Generator.

4. Reinforcement Learning: Refine those responses over time based on the analysis of how well the Conversational AI did this go-around.

According to various viewpoints, conversational AI systems can be grouped. Previous research has classified these systems into categories such as text-based and spoken dialogue systems, voice user interfaces, chatbots, embodied conversational agents, robots, and placed agents. Conversational AI systems are classified based on their methodologies, channels, and goals. The many classification approaches show how quickly the domain has developed and the amount of attention it has received[14].

Different companies have been employing their research and development teams to explore potentialities and create products based on this technology. Nowadays, one of the most important companies has become an exponent when it comes to AI systems. OpenAI is a research lab that has a focus on using artificial intelligence for the benefit of humanity. OpenAI firstly launched ChatGPT using the GPT 3.5 model, and later on the GPT 4.0 model. Both models are a benchmark regarding AI models since they are more capable compared to other models in the market.

2.4.3 Advantages and disadvantages related to the technology

What makes AI performance superior to that of humans is the fact that AI bases its decisions on facts rather than emotional factors, and this is where they gain such supremacy in performance. However, AI systems have many disadvantages when used.

Table 2.1 describes some of the advantages and disadvantages of artificial intelligence:

ADVANTAGES	DISADVANTAGES
Enhanced Productivity	Impacted Labour Market
Efficient Parallelization of Tasks	Absence of Human Emotional Engagement
High Success Rate	Substantial Resource and Time Expenditure
Reduced Error and Defect Rates	High Development Costs
Mitigation of Human Error	Limiting Creative Problem-Solving Abilities
Autonomous Risk Management	High Unemployment Rates
24/7 Availability	Technological Dependency
Expedited Decision Making	Potential for Reduced Motivation in the Younger Generation
Pioneering Innovations	

Table 2.1. Advantages and disadvantages of AI

The use of artificial intelligence software/tools has several advantages. It can be said that the biggest advantage these systems bring is the fact that they rely on factual information-based systems for decision-making tasks, rather than emotions that could, at times, bring entropy into human judgement. That is to say, in terms of repetitive, routine, and automated tasks, AI may, in fact, be much better than humans at performing them, as it has large processing capabilities with low error rates.

The other aspect is that these systems can make predictions starting from an analysis of data trends, sometimes providing crucial insights into some critical problems and scenarios. This characteristic determines the speed and precision of these instruments. This, of course, means that these attributes of AI can make it perform better in parallel tasks, thus improving productivity, effectiveness and much lower turnaround times for tasks compared to tasks done by humans. The AI systems themselves can be everywhere, taking up as little space as possible in the real world to make them portable and usable[15].

However, artificial intelligence also has several weaknesses. This happens especially when these tools are not used correctly and, if used improperly, they could provide incorrect information on information sources and facts, thus causing damage to various aspects.

The more it becomes an integral part of life, the more it continues to create dependency and laziness in the younger generation, who can use these systems as shortcuts to every productivity task without making any effort to do so. It imposes human dependence in making decisions or carrying out tasks independently and even blocks creativity. It is dangerous in a way that creates dependency on technology not controlled by end users and could likely lead to technological misuse of some sort of misinformation with unintended consequences.

Furthermore, AI systems do not possess emotional intelligence nor do they share the morally and ethically right perspectives of humans.

Users should be aware of the powerful opportunities and serious drawbacks of artificial intelligence that consciously emerge with respect to some critical points of these systems. This could help users use AI responsibly, while being mindful of the harm it could create[15].

2.4.4 Ethical Concerns Related to AI

Artificial intelligence is developing and progressing rapidly. That means users use this technology in different aspects of daily life to facilitate their daily tasks. Since its effect on productivity is evident, people are using it exponentially. For this reason, specific ethical concerns arise about its usage. The main shortcoming is the confusing legislation, regulation and discussion to discuss its responsible use and impact on the society[16, 17].

Figure 2.9 above shows several ethical concerns of AI that arise in different aspects of life. One of the main characteristics defining artificial intelligence's nature is that it is trained on large data sets. From this perspective, the solicitude of algorithmic bias derives. It refers to the possibility of these systems making judgements or making decisions based on the data they are trained in, which is only sometimes the most exact answer and, in contrast, might give discriminatory results based on critical factors such as race, gender, or socioeconomic status, when used in fields such as employment, financial

Figure 2.9. Ethical Concerns of AI technology

systems or judicial systems.

An account must be created to use AI tools or software. This creates privacy and data protection concerns because it relies on large amounts of personal data. More transparency should be provided regarding collecting, storing, and using data. Transparency constitutes another ethical problem since AI's nature is challenging to decipher. This makes it difficult to understand the reasoning or logic behind their results, decreasing their credibility and accountability. Accountability and liability issues worsen as well due to assigning or responsibility: for example, in cases of damages or errors produced by the system, the percentage of responsibility shared by stakeholders, users, or the AI system itself is unclear.

Moreover, in recent years, the number of people exploring AI for personal uses and as a supportive tool at work has increased. Job displacement and economic inequality concerns arise from AI's fast integration into the work environment, especially regarding automatic tasks. This brings worries about job displacement and changes in the labour market, affecting society.

Another critical aspect is AI's ability to manipulate and misinform society when used irresponsibly. These tools, whether used intentionally or unintentionally to create disinformation or share news not verified by the source, might destabilise. It may lead to manipulating public opinion and the society's overall stability. That is why it is essential to design AI systems in a fair, equal and impartial way.

Furthermore, another ethical concern is AI's environmental impact, which comes from the fact that the infrastructure used to power these systems has a significant energy consumption and carbon footprint. Analysing this effect is crucial to understand these systems' longevity from an environmental sustainability viewpoint.

In conclusion, these are only some of the ethical concerns and their explanations related to AI systems. Open discussions with experts in the field, regulations, and legislation are important to minimise the risks that irresponsible use of AI systems might bring in different areas of life[16, 17].

Chapter 3

AI Applications

3.1 AI applications in Education and Academia

Artificial intelligence technology has started to be used massively in different sectors. It has the potential to bring innovation and transformation in the way business-related processes were previously managed.

Education is one sector where this technology can bring innovation. In this environment, AI can do this by revolutionising teaching and learning methods and the administrative processes that comprise them [18]. The combination of AI in educational systems has given the first benefits of a personalised learning experience, greater engagement of students and saving time devoted to routine administrative processes[19]. Moreover, integrating AI into higher education can also bring innovation to teaching and learning. We can thus mention automated grading systems and personalised learning platforms that use this technology[20].

However, using this technology in the educational environment also brings challenges and ethical concerns. Thus, policymakers and educators must ensure that students have all the necessary knowledge to use this technology responsibly. Some of these suggestions may include understanding the basic algorithms that AI uses, critically evaluating AI applications, and considering the ethical implications that the usage of AI might bring when used in different contexts.

The impact of using artificial intelligence in academia has several implications [19]. First of all, this technology can benefit the teaching and learning process since it can offer personalised learning experiences, adapting to each student's personal and specific needs. Also, it can introduce new teaching methodologies that would not have been possible doing before[21]. For example, AI could help customise learning based on the strengths and weaknesses of each student. On the other hand, educators and teachers could benefit from the automation of routine administrative tasks, thus having more time to focus on educational activities that are more important. Also, academic applications based on AI technologies have brought considerable innovation in the field of researching scientific articles, analysing data, and finding articles related to the subject under review. This feature makes knowledge generated more easily and information easily accessible. This increases the possibility that the scientific results of the work are more accurate and efficient and also facilitates the work of researchers. If before, a researcher or professor dealing with scientific research had to read dozens of articles that were probably useless in their work, now, different AI software can proofread the article and give a framework of what is addressed in the article, helping them to stay updated on the latest developments and consider studying articles that are relevant to their work. The challenges and opportunities of using AI in universities are extensive[20].

As mentioned, the biggest and most important challenge involves addressing data security and privacy concerns, especially when the latter is used to collect and analyse student information. Another challenge would be to provide this technology equally for all students without promoting a digital divide.

Finally, it is important to understand the algorithms used in educational decision-making processes, thus increasing the control of educators and professors over this technology. To face these challenges, academic curricula must prioritise the theme of incorporating AI in the educational environment responsibly[22] . There are premises that the future trends of using AI in academic curricula and the teaching process will develop even further with the development of this technology. What is worth emphasising is the fact that there is a need for empirical studies to really measure the effect and indication that the approach of this technology can bring to this field by comparing it with traditional methods[23].

3.2 AI applications in Industry

The development of artificial intelligence portends an epochal change in the automotive sector. Artificial intelligence has enormous potential to change diverse areas of the automotive industry, from manufacturing and the assembly process, to vehicle performance and its safety features [24]. Artificial intelligence can help automakers develop vehicles that are more effective and less expensive in their production[25].

The other large sector where artificial intelligence is helping a lot is the development of self-driving vehicles. These autonomous cars are developed by artificial intelligence algorithms which, in turn, interpret data from sensors, cameras and other components to develop their way of driving and make the right decision at the same time. Other areas applied to artificial intelligence include predictive maintenance, whereby artificial intelligence technologies are applied to the analysis of information from vehicles to identify a possible problem long before it occurs, thus reducing downtime and increasing overall reliability[33].

Petrol and oil companies are also using artificial intelligence in their operations. The oil and gas industry is increasingly turning to artificial intelligence, promising both better decision-making and better operational efficiency. Artificial intelligence technologies allow companies to analyze a large amount of information, in the area of exploration, drilling and production, refining process. This enables precise models to make operational improvements and better prediction of phenomena, such as oil reserves or the amount produced[27, 28]. The aforementioned technology is also implemented in the automation and optimization of refinery operations, from monitoring and controlling crude oil flow to inventory levels to predicting equipment maintenance requirements, among other uses. Artificial intelligence can also be used effectively to predict or identify potential dangers and security risks. For example, AI-based real-time sensor data monitoring and analysis of pipeline leaks or equipment failures[29].

Moreover, the use of artificial intelligence can support the advent of materials science. Customised AI algorithms, by contrast, can easily sort through massive data sets looking for new materials that possess the necessary properties. In the context of synthesis and processing, the outlined approach provides the opportunity to improve and predict methods of materials behaviour under different conditions. This, in turn, not only makes it a process of discovery and development, but also enables the creation of innovative high-performance materials that were not available, as easily perceived previously. AI can also help optimise better-designed chemical processes to increase efficiency and reduce costs. It is said that the chemical industry will revolutionise with the use of artificial intelligence in many of its aspects. For example, artificial intelligence could be useful for optimising the formulation of chemical products. For example, it could predict the outcome of a reaction and/or predict any risks or side effects resulting from the use of that chemical[30].

AI can also be useful in the development of new catalysts for chemical reactions, therefore increasing the pace of research and, at the end of the day, allowing to set up a more sustainable and ecologically viable process. Across different industries, AI has proven to bring numerous benefits to the automotive, petroleum, material, and chemical sectors. The same includes the improvement of the efficiency of the manufacturing processes, the saving of costs due to the process of automation, better targeting of marketing activities, betterment of supply chain management. Just like any new technologies, the industrial applications are not a matter of exemption when it comes to challenges arising in other normal technologies[31].

Most of the challenges associated with the integration of AI into industrial applications take several forms [24], with the most relevant regarding data availability and quality. For AI to be effective in analysis and providing insights, the same will require lots of but structured and quality data. However, it means that data could not be in many industrial setups, whereby the same is not structured, disaggregated, and mostly not complete to enable functioning AI systems. Firms would need to invest in data gathering and management processes, ensuring that the clean and reliable data exist[32].

Examples of the best success in implementing AI in the automotive industry include the invention of the self-driving car. These are means of transport bestowed with abilities for them to make decisions while in transit by incorporating technologies such as AI, computer vision, and machine learning algorithms in their decision-making. Petroleum is one more area of business where the AI successfully found its application. It found its application to optimise the process of drilling, forecast the behaviour of the reservoir, and so on[33].

3.3 AI applications in Finance

Technology has provided the influence through which the business practices involved in the financial services industry have been totally revolutionised [34]. The modern banking system has undergone numerous changes compared to the traditional financial system, which was characterised by numerous manual activities. This is because they use various technologies: artificial intelligence, cloud service, virtual reality, augmented reality and now climate sustainability [35].

These technological advances have influenced not only the work process within financial institutions but also how customers perceived these institutions. In practical terms, there are large financial companies that have invested in providing a superior customer experience through tracking, personalisation and optimization of the customer journey [36]. One of them is the emergence of robo-advisors in the field of financial technology with the use of artificial intelligence, which is capable of engaging customers extensively [37].

Artificial intelligence technology applied in data analysis has brought great benefits to organisations dealing with large financial data. The application of artificial intelligence in financial data analysis has greatly redefined the way in which financial institutions undertake the process of analysing and interpreting them. In fact, AI tools can process large volumes of financial information faster than human ability can, providing greater precision in analysing financial information while saving time. AI can also track patterns and trends in financial data that humans might not notice, ensuring that financial institutions make the best decisions and predictions[38].

The impact of artificial intelligence in the financial environments of institutions has been radical through the generalisation of financial decision making. First, AI algorithms can examine large volumes of financial data for patterns or anomalies that would influence investment decisions. This can, therefore, enable financial institutions to make better, data-driven investment decisions that are less risky than errors or biases made by human decisions. This informs the credit waiver decision which can, in turn, lead to a reduction in instances of default risk and improved portfolio performance[38, 39].

One of the main areas where artificial intelligence has influenced finance is risk management. The ability of artificial intelligence to analyse big data and extract it, or in some cases, even identify risks or deviations, is revolutionary for the world of financial risk. This helps financial institutions to properly assess and manage different types of risk, which include, among others, credit risk, market risk, operational risk and fraud. Artificial intelligence techniques could allow banks to build much more complicated models that would use many more variables, many more than would ever be possible with manual analysis. This could really change the operational reality by isolating problems with a much greater degree of precision and provide insights into what the most effective ways to mitigate these risks are[40].

For example, a case study focused on how AI technologies are implemented in financial sector is portfolio management, whereby artificial intelligence analyses large volumes of financial data to help optimise the investment portfolio against established risk and return objectives[40].

Another area where artificial intelligence can be applied in the financial sector concerns the detection and prevention of fraud. For example, it may be that transaction data is viewed by AI algorithms to spot suspicious patterns or anomalies to detect fraudolent activities[38].

Overall, the application of artificial intelligence in the financial sector can go a long way in improving decision making, more appropriate risk management and better fraud detection. The scope of artificial intelligence in the financial sector is broad and holds great promise[40].

In summary, the arguments listed before mean that AI can boost the financial sector in the ways described: from better, faster and more precise data analysis to improving the practice of risk management and automating different processes. Therefore, the use of artificial intelligence and machine learning in the financial sector has proven to be a saviour of decision making, risk management and the entire financial services automation process[41].

3.4 AI applications in Medicine and Healthcare

Artificial intelligence is trying to create new paradigms in the fields of health care and medicine by developing software and tools that replicate human cognitive functions. This technology has the potential to revolutionise health care in several aspects: increasing patient safety, improving the efficiency of administrative tasks, and improving predictive outcomes[42, 43].

Furthermore, another characteristic of these developments is that AI can make predictions and categorisations related to the risk of diseases by identifying patients who may be more exposed to the risk of developing certain diseases or adverse events and, therefore, provide guidance to clinicians to address interventions and analyse data efficiently. Also, AI brings greater efficiency to patients and doctors by automating routine tasks, such as administration and data entry. This allows healthcare workers to spend more time directly on patient care. Applying artificial intelligence to improve operations within the healthcare organisation also reduces unnecessary costs[42, 44].

Moreover, the vast dataset analysed with artificial intelligence technologies may reveal hidden patterns or trends that human doctors may not be able to perceive quickly enough. This can assist doctors in the fastest diagnosis of diseases, and therefore helps in the most efficient treatment with patient results[42, 43].

Artificial intelligence can also help in prevention by using patient information to predict all possible health risks that have not yet appeared, thus helping to prevent various health conditions. This functionality can help physicians translate knowledge into evidence-based clinical decision-making and healthcare delivery in a value-based manner to achieve better overall quality[45]. Overall, in summary, AI is revolutionising the healthcare sector in four key macro areas:

- Patient care
- Research
- Administrative duties
- Medical Education

Based on machine learning algorithms, the above AI-based diagnostic approaches and tools offer good prospects for several healthcare fields such as: accurate diagnosis, precision treatment, drug development, virtual clinical practice, disease diagnosis, prognosis and drug management, health management and monitoring and the like. These advances have the potential to provide personalised healthcare for patients. However, such potential benefits come with ethical considerations and challenges[46]. The application of artificial intelligence in health care involves taking into consideration the security of patient data as the latter are considered sensitive data from a security point of view. Another disadvantage of the AI system is that it cannot reason with "common sense" or "intuition and clinical experience" as human doctors do[42].

Chapter 4

Methodology

4.1 Introduction

This thesis aims to provide an overview of how efficient, correct, and promising methods based on artificial intelligence are intended to offer a way to automatically generate diagrams for process modelling. This assessment is done through a comparison between the respective solutions of BPMN diagrams developed by humans and AI models, in particular GPT-3.5 and GPT-4.0.

This work therefore establishes an important baseline by placing importance on developing capabilities in the technical and creative domain for AI by closely comparing the results of AI-generated outputs with human-created solutions.

The strategy was sequential through iterative understanding and development rather than simple comparison. These were part of improving the input requests of this study, understanding that early AI trials can sometimes be mediocre, but their effective deployment is part of achieving maximum AI efficiency. This iterative process improved the quality of the diagrams produced by the AI and provided deeper insight into the specifics of how the AI learns and can be adaptive in technical jobs.

4.2 Background and Literature Review

As anticipated in previous chapters, BPMN diagrams play a fundamental role in communicating business processes within an organisation. Business analysts primarily use it to communicate business processes to developers and stakeholders, trying to ensure a better and easier understanding of how the processes work. Evaluation criteria for BPMN diagrams are important as they help ensure the quality and effectiveness of these diagrams. The open research question would be how to evaluate the quality and effectiveness of BPMN diagrams in practice[47], and to have the ability to demonstrate that two structurally different diagrams are behaviorally the same.

Therefore, to answer this question, experts and professionals in the field in question have proposed different evaluation criteria for BPMN diagrams from the point of view of clarity, effectiveness, correctness and completeness. This section will discuss several ways to evaluate BPMN diagrams.

One approach employed to measure the equivalence of diagrams would be through mathematical techniques. These techniques can be used even if the diagrams examined may have structural differences but also to analyse the underlying behaviour and logic of BPMN diagrams. Their purpose is mainly to determine whether two or more diagrams are functionally equivalent, regardless of their structural differences[48]. These techniques include:

1. The definition of the evaluation criteria
2. The establishment of formal relationships
3. The verification of the transitivity and equivalence properties
4. The equivalence check

In order for two processes to be considered equivalent it becomes necessary to outline the specific conditions that must be satisfied. This is done by providing mathematical definitions for each type of equivalence and describing structural and behavioural details. Mathematical notations and structures are essential tools in representing Business Process Model and Notation (BPMN) processes. Using mathematical models it is possible to make accurate and precise evaluations of processes, ensuring efficient and effective implementation. The verification of transitivity and equivalence properties is necessary to guarantee the coherence and reliability of BPMN diagrams [49].

In these steps, mathematical proofs are used to formally demonstrate the transitivity and equivalence properties of these diagrams. In addition, several mathematical methods, such as automata theory, process algebra, and Petri Nets, can be used to check the equivalence of BPMN diagrams and process equivalence checking. This helps to compare different process representations and systematically compare them to identify any inconsistencies or differences in their behaviour[49].

Another way to evaluate BPMN diagrams is to evaluate their appropriateness. This occurs through the workflow process, which is composed of control flow, data flow and resource capabilities. This type of analysis aims to highlight the strengths of BPMN, the weaknesses and the adequacy of BPMN, thus ensuring a complete evaluation.

The *Workflow Pattern* is a set of general, repeatable constructs that were originally created to evaluate workflow systems but can now also be applied to evaluate workflow standards and business process languages. In these constructs three different perspectives are presented.

The first is the control flow perspective, which makes a general evaluation of the key parts of the diagram, such as tasks, gateways, order of events. Through this model it is possible to evaluate how precisely the typical parts of the control flow can be described, thus evaluating the effectiveness of the process modelling [50].

Another perspective that can be distinguished is the data perspective, which, as can be deduced from the name itself, has to do with the construction of data. These constructs are classified into different models, which include data visibility and data transfer. Using this perspective to evaluate process models helps you better understand the definition and use of data in a process context.

Lastly, the resource perspective focuses on evaluating aspects of resources, which include elements like pools and lanes. In this way, it is assessed whether resource-related characteristics are reflected well in process modelling diagrams [50].

4.3 Study Design

The core of this research consists of a set of exam papers taken from the final exams of the Information Systems course part of the Master's degree in Management Engineering at Politecnico di Torino, covering the years 2017 through 2022. These exercises, which range in difficulty and necessitate solutions for both BPMN diagrams, were carefully selected to provide a thorough assessment. The final papers were supplied in a ZIP file by the course teacher and provide a solid foundation for examining modeling capabilities in both AI models and human specialists.

The study started with a systematic approach to tackle each exercise. It began by focusing on BPMN diagrams and identifying crucial elements such as processes, gateways, flows, pools, and lanes. The tool utilized for BPMN diagram construction during this phase was BPMN.io. The process started with a thorough understanding and modeling phase and gradually became more efficient and proficient in diagram construction.

The exercises were solved using advanced GPT-3.5 and GPT-4.0 AI models that have highly sophisticated natural language processing and generation capabilities. To do this, each exercise was inserted into these AI systems using standardized prompts that closely resembled the instructions and information provided during the human exercise-solving phase. The resulting AI-generated outputs were then gathered and organized for a thorough comparison with the solutions generated by humans.

During the study, an iterative refinement procedure was applied, particularly for the AI-generated solutions, to accommodate the possibility of early errors or defects. This involved adjusting the input prompts based on preliminary outputs, expanding the context provided, and refining instructions to steer the AI towards generating more accurate and contextually relevant diagrams. Each iteration was meticulously documented, recording the alterations made and their impact on the quality of the AI-generated diagrams. This procedure was critical in assessing the AI's adaptability and learning progress in challenging diagram production tasks. This thesis aims to compare human and artificial intelligence (AI) capabilities in generating BPMN (Business Process Model and Notation) diagrams. The purpose is to highlight the potential efficiencies and challenges associated with using AI for complex diagram generation tasks. It also shows the iterative nature of obtaining refined AI-generated outputs by detailing each phase, ranging from exercise selection to final diagram generation and iterative refinement.

To conclude, there are four main steps (presented in Figure 4.1) to follow when tackling an exercise: exercise selection and preparation, human solution generation, AI solution generation, and iterative refinement. These steps involve carefully selecting and preparing for the exercise, coming up with a solution through human and AI input, and refining the solution through a collaborative iterative process.

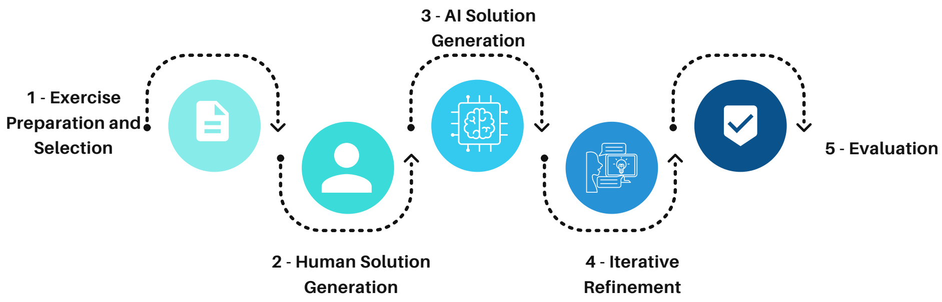

Figure 4.1. Evaluation Process

4.4 Prompt Design

ChatGPT and LLM distinguish themselves by being able to generate responses in the manner of a human. Another aspect is the model's ability to respond to subsequent queries while preserving a logical flow when new requests are introduced. Regardless of this characteristic, the most difficult task was directing the model to provide the correct answer based on the inquiry, rather than an answer that was, in technical terms, hallucinative.

This section will describe how, after multiple efforts, it was feasible to identify the right question to obtain the correct response. Firstly, it is necessary to explain the idea of *Prompt Engineering* in order to comprehend this reasoning. Prompt Engineering consists of a collection of methods and strategies that are used to design, write, and optimise the commands provided to the LLM, or "prompts," so that the user receives an answer from the model that is accurate not only in terms of the construct but also in terms of factual, concrete, and repeatable content. Experts in the industry have attempted to establish various approaches by considering the aforementioned facts in order to produce a prompt that is as precise as possible[51, 52].

In the context of this study, the goal of prompt design was to repeatedly try and adjust the prompt so that the response provided was the most pertinent one. First, the goal was for GPT 3.5 and GPT 4.0 to be able to construct the BPMN diagram using the BPMN 2.0 formalism, given the primary problem and the graphical specifics. Thus, in addition to testing the model's logical analysis, the capacity of the latter to produce the final diagram with a graphic representation via an XML file was also sought after.

The GPT 3.5 and GPT 4.0 models' databases are trained in the internet database. Having stated that, one job of the prompt was to direct the database search for knowledge specific to the subject. This is a rule that ought to apply when the response to be gleaned by probing the model is likewise domain-specific. Conversely, the response that was given would be untrue. The prompts also need to follow the proper format, be precise, explicit, and use clear language. In reference to the final principle, examples may be included in the prompts if needed.

The prompts can be classified as few-shot or zero-shot prompts depending on how many examples are provided to the models from the initial query. Few-shot prompts require some examples in order to provide a more accurate and helpful response, while, on the other hand, zero-shot prompts do not require any examples.

Many experts in the field agree on the fact that prompt perfection is a repeated process. AI models benefit from all of the previously described ideas since they provide a distinct domain context. Artificial intelligence is able to produce a more precise, tangible, and specialised response based on the knowledge domain by better comprehending the user's input and the inquiry's goal[52].

4.4.1 Prompt Development Process for BPMN Diagrams

The so-called GPEI methodology is one of the approaches employed in this study. The GPEI methodology stands for Goal Prompt Evaluation Iteration. It is a four-step process that includes defining a goal, designing the prompt, evaluating the answer, and iterating to achieve an adequate response[52].

As previously stated, the process of fine-tuning the prompt is iterative, even within the particular context of the study under consideration. Observing how the responses produced by GPT 4.0 and GPT 3.5 would alter as the prompt changed dynamically and how its capacity to produce responses altered is another factor, in addition to the previously mentioned ones. It is also fascinating to discover how, as the model learns more from the first prompt, the final response becomes more specific.

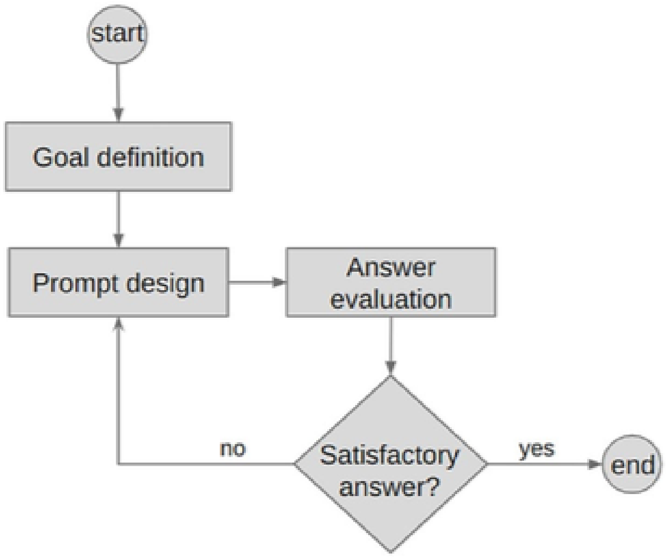

Figure 4.2. GPEI Methodology, Source:[52]

Figure 4.3 shows the initial prompt used to interrogate AI models:

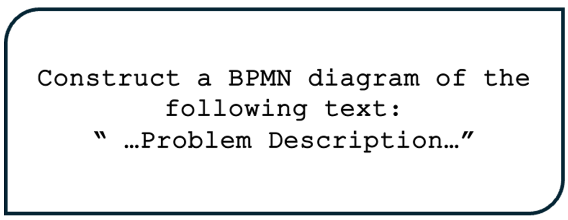

Figure 4.3. Initial Prompt

This question contains many mistakes in itself, being that it is very vague and the question leaves room for guesswork. The response of GPT 3.5 was a textual representation in which only a logical flow of the problem could be distinguished, but nothing more. This is because the question did not contain anything more specific.

As can be seen from Figure 4.4, the initial prompt failed to provide reliable results. For this reason, the second attempt was to transform the above format into XML format (being that the BPMN diagram modeling tool accepts the BPMN format). In this part, both the GPT 3.5 model and the GPT 4.0 model were not able to create a complete XML file, leading to unsuccessful

```
Start
  |
  V
Compile form
  |
  V
Load documents and information
  |
  V
Verify data
```

Figure 4.4. Textual Representation of solutions generated by GPT 3.5

file uploads.

Another strategy that was followed was as follows:

1. Feed AI with the format of a BPMN diagram
2. Feed Problem Description
3. Solution Generation

The GPT 3.5 and GPT 4.0 models performed slightly better in this case, as evidenced by the ability to view the diagram during file loading, despite the gaps and non-continuous sections that separated the sections. Separating the components of the answer was another approach. Since there was a limit to the code generation, the GPT 3.5 and 4.0 models were asked to divide the answer into multiple parts. However, even in this situation, neither model was able to produce consistent answers between them, indicating a limit of the GPT models concerning graphical representation.

The goal of every stage, from creating the prompt to improving it, is to make sure that all of the responses are consistent. The foregoing attempts were utterly ineffective and produced no reliable outcomes, thus a full strategy change was necessary. Another tactic was to ask the model to identify pools, lanes, processes, and gateways logically and descriptively, and then to have these diagrams built by hand using the guidelines provided by the GPT 3.5 and 4.0 models, given the obvious limitations of these models in creating diagrams in XML format.

Figure 4.5 summarizes the different steps used to refine the prompt:

After having analysed and fine-tuned the prompt engineering process, the last prompt is the following:

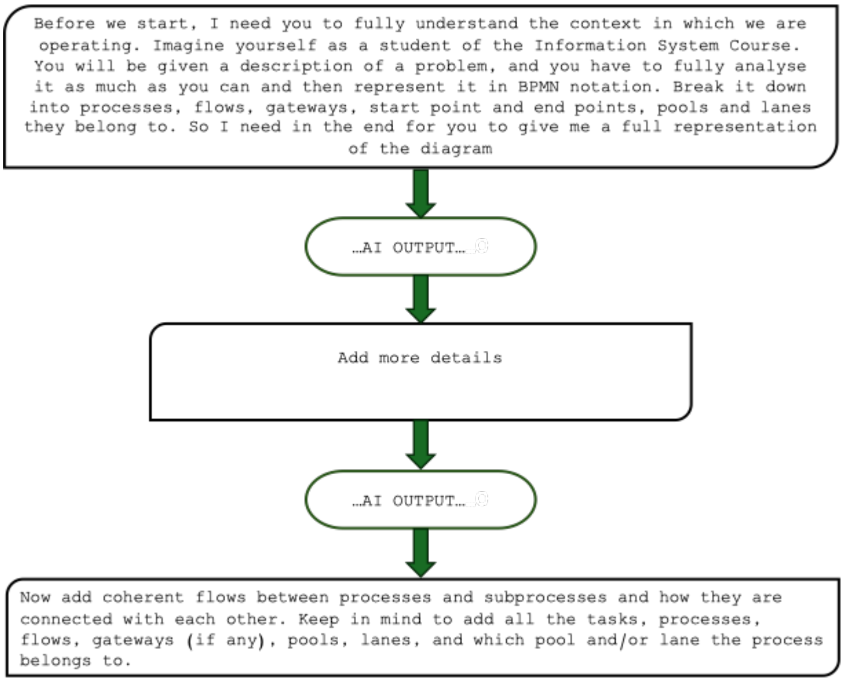

Figure 4.5. Prompt Refinement

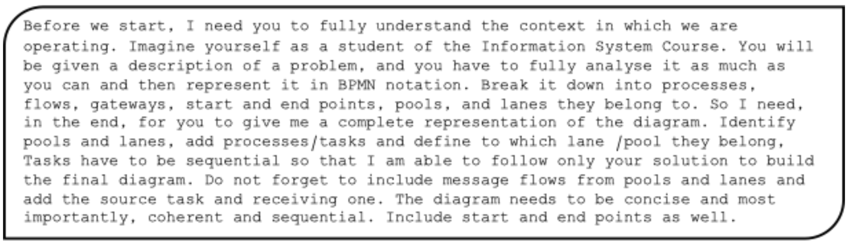

Figure 4.6. Final Prompt

This prompt was common to both GPT 3.5 and GPT 4.0 models. The reason for this choice is the fact that one of the objectives of the thesis was to analyse the limits of the two models. Starting from the same prompt, the goal was to assess how do the capacity and ability of the two models to generate correct answers differ, and what would the differences be between human-generated and AI-generated diagrams. A preliminary expected first result was that, starting from the same data and prompt, GPT 4.0 would normally have better performance in generating diagrams in comparison to GPT 3.5.

Regardless of the final prompt, based on the complexity of the problem description, AI was not always able to give a final answer. When the problem was very complex, there was a need for some integrative information in the form of instructions.

What was expected from the AI models as a response to the prompt was to be able to create a logical sequence of all the included events, starting from the start point, processes, gateways, pools, lanes, message flows, and endpoints. Unlike GPT 3.5, GPT 4.0 showed a more specific knowledge domain regarding the BPMN 2.0 formalism because, among the events, it managed to connect the icon that had to be placed in the diagrams.

The hypothesis of this study was the fact that GPT 4.0 should perform better than GPT 3.5 in creating the logical sequence and, as a consequence, give a more accurate and complete diagram than GPT 3.5, taking into account that it is a model trained on a larger amount of data. This was the reason why the same prompt was used in both models: to test the hypothesis that, starting from the same prompt, GPT 4.0 can create a better response than GPT 3.5 and manage to generate diagrams that are comparable to human-generated ones, and to assess whether AI can generate or assist humans in generating BPMN diagrams.

4.5 Evaluation Methodology

4.5.1 Evaluation Criteria

An important part of the study in question was also the methodology for the evaluation and interpretation of the results. After refining the final prompt, for each exam paper, the GPT 3.5 and GPT 4.0 models were able to generate a logical sequence of all the constituent components of the diagram, and then these diagrams were constructed by the student conducting this thesis following these instructions.

Once all diagrams were completed, there was a need to evaluate, for each exam paper, the human-built diagram, the GPT 3.5-built diagram, and the GPT 4.0-built diagram to compare them with the solutions.

The final objective of this thesis was to verify if AI models were able to create comparable results and, if so, if they managed to outperform human-generated diagrams.

To carry out this evaluation, it was necessary to determine the evaluation criteria. In this case, the evaluation criteria were focused on two specific groups:

1. Syntactic Evaluation
2. Appropriateness Evaluation

Syntactic evaluation deals with the accuracy of the symbols used in the diagrams generated by the three actors about the symbols, language, and syntax of the BPMN 2.0 formalism. This assessment is also the simplest to make from a practical point of view since these errors are observable more objectively and rationally.

As mentioned in the previous sections, for example, from the behavioural side, two diagrams can have different structures but be correct, while in this case,part of the evaluation is only the diagram's accordance with the symbols and the language of the formalism.

Another criterion for evaluating the diagrams was appropriateness. This was done through the WorkFlow Pattern (which was explained in section 4.2), and the evaluation was based on the perspective of control flow and resource flow. In this way, a general comparative evaluation of the basic parts of diagrams, pools, and lanes as resources was made.

The reason why only these criteria were used and not other criteria such as practical and semantic was because, firstly, being exam papers for academic use and not necessarily presenting a genuine business project, it is difficult to understand from a comparative perspective if a diagram is more accurate from the semantic point of view. Another reason is the lack of tools and professional expertise to perform such an analysis.

4.5.2 Evaluation Process

The evaluation process was built up in several specific steps:

1. Creating a directory for each problem with the solution provided by the professor and the solutions from the three actors (human, GPT 3.5 and GPT 4.0)
2. Definition of the scoring system
3. Evaluation of appropriateness of all diagrams, for each exercise and each actor
4. Syntactit evaluation of all diagrams
5. Construction of the matrix of final results, one for the assessment of appropriateness and one for the assessment of syntax errors.

The first step was the simplest. For simplicity and practicality reasons, different folders had to be created and divided according to the number of exam papers, each of which contained the solution given by the professor of the Information Systems course and the solution to the exercises by each actor included in this study.

The second step consisted of defining the scoring system regarding the evaluation of the appropriateness of each diagram. This was done by agreement, taking into account all cases in the study. First of all, the assessment of appropriateness was carried out from two perspectives:

1. Resource perspective: in this section, it is verified if the pools and lanes defined in the solutions of the exercises are coherent with the real solution of the exercise

2. The perspective of control flow: in this section, it is verified if the basic parts of the diagram (such as start events, tasks, gateways, timer events, and end events) are present in the solutions, as well as from the logical side if the order of presentation of the events conforms to the description of the initial problem.

A matrix containing the main information was built as follows:

As can be clearly seen from the above table, the "conditions" column is composed of all the main parts of the problem solution. In this way, by means of a scoring system, it is determined whether the human-created diagram is coherent and contains these parts or not.

After some reflection, the scoring system was decided as follows:

'1' - Correct
'0.5': Partially correct
'0': Incorrect or inexistent

Table 4.1. Scoring System

Thus, three different cases can be distinguished:

- Index 1 means that the component in the diagram solved by one of the actors (human, GPT 3.5 and GPT 4.0) is the same as in the correct solution

- Index 0.5 means that the component in the diagram solved by one of the actors (human, GPT 3.5 and GPT 4.0) is not the same as in the correct

NO. CONDITION	SOLUTION
POOLS	
1	Pool 1
LANES	
2	Lane 1
3	Lane 2
PROCESS PARTS(EVENTS)	
4	Process part 1
5	Process part 2
6	Process part 3
ORDER OF EVENTS	
7	Process part 1 --> Process part 2
8	Process part 2 --> Process part 3

Figure 4.7. Matrix of Evaluation Criteria

solution but is close enough, by having a partially correct definition or by belonging to a different pool

- Index 0 means that the component in the diagram solved by one of the actors (human, GPT 3.5 and GPT 4.0) is not the same as in the correct solution, in the sense that it is either incorrect or it does not exist at all

The final matrix takes this form:

For each condition, in each row and in each column as an actor, an index between 1, 0.5, or 0 was determined. This was done repeatedly for the entire exercise set, consisting of 19 exercises.

On the other hand, regarding syntax evaluation, the approach was very straightforward. The evaluation process was done with a tool provided by the Department of Computer Engineering, and in this way, the syntax problems for each diagram were quantitatively detected.

The syntax errors that were evaluated are presented in Table 4.2:

The results are reported in the format show in Table 4.5.2 and were repeated for all 19 exercises:

NO. CONDITION	SOLUTION	HUMAN	GPT 3.5	GPT 4.0
POOLS				
1	Pool 1			
LANES				
2	Lane 1			
3	Lane 2			
PROCESS PARTS(EVENTS)				
4	Process part 1			
5	Process part 2			
6	Process part 3			
ORDER OF EVENTS				
7	Process part 1 --> Process part 2			
8	Process part 2 --> Process part 3			

Figure 4.8. Final Matrix

Error Type	Error Description
No implicit split	A task having more than one outgoing flow
superfluous-gateway	A gateway having only one outgoing and one incoming flow
single-start-event	A pool/process having more than one start event
end-event-required	A pool/process having zero end events
start-event-required	A pool/process having zero start events
no-implicit-end	An element not having an outgoing sequence flow that eventually leads to an end event
no-implicit-start	An element not having an incoming sequence flow that has the start event as root
no-duplicate-sequence-flows	A pair of elements being connected by multiple sequence flows
fake-join	A task having more than one incoming sequence flow
no-gateway-join-fork	An exclusive gateway having multiple incoming and outgoing flows at the same time
boundary-event-task-flow	A task with a boundary event not having a regular outgoing sequence flow
no-disconnected	An element having no sequence flow
no-incorrect-outgoing-message-flow	A message flow starting from an element that cannot send messages (e.g. a user task)
no-bpmndi	An element missing its visual representation information in the XML structure

Table 4.2. Syntax Errors

Exam paper	Actor	Syntax Errors
Exam Paper No. 1	Human	Number of human syntax errors
Exam Paper No. 1	GPT 3.5	Number of GPT 3.5 syntax errors
Exam Paper No. 1	GPT 4.0	Number of GPT 4.0 syntax errors

Table 4.3. Reporting format for syntax errors found in the diagrams

To assess the suitability of a small number of exercises, the same student who solved the problems carried out the evaluation. Firstly, the candidate solved the exercises, then the BPMN diagrams were solved by GPT 3.5 and GPT 4.0, uploaded to the BPMN.io application, and evaluated by the same student. The appropriateness evaluation was manual and was recorded on a sheet, while the syntactic evaluation was automated.

4.5.3 Sample Evaluation

This section will present a few examples of evaluated diagrams, to understand how the evaluation criteria were applied in this particular case.

The following problem requires building a BPMN diagram to build an IT system that helps travelers, tourists, and healthcare institutions manage testing during the pandemic. The system, using the personal data recorded by travelers/tourists, records the destination and method of travel and generates a unique code. The traveler/tourist must communicate this code upon arrival at the health facility and undergo the rapid test. On the other hand, the system also offers the functionality of recording the test result. Let's see concretely how the three actresses behaved in generating the diagrams.

In the following case, it can be seen that the appropriateness evaluation criteria have been applied by assessing pools and lanes as elements as the first step. It can be observed that there are two pools in this problem: Travel and User Registration, as well as three lanes: User Registration, Employee and Registered User. In Figure 4.9, the solution to the problem was compared to the student's solution.

Considering the differences between the two variants, the evaluation accuracy scores were assigned as shown in Figure 4.10:

Overall, considering only pools and lanes as elements, the human-generated diagram is correct, and therefore, when the conditions are correct, the index 1 (i.e., 100% correct) is assigned. The only difference is between the trip pool (provided by the solution) and the route pool (provided by human-generated

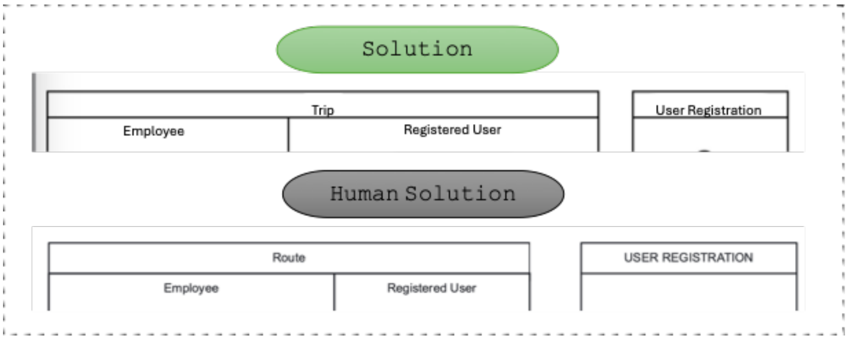

Figure 4.9. Solution Vs. Human

NO. CONDITION	SOLUTION	HUMAN
POOLS		
1	User Registration	1
2	Trip	0.5
LANES		
3	New user	1
4	Registered user	1
5	Employee	1

Figure 4.10. Evaluation Scores Human vs. Solutions

solution). These notions, in fact, are not entirely equivalent to each other, and since the definition can be misleading, it is therefore partially correct. For this reason, it was assigned 0.5 points.

The solutions performed by the GPT 3.5 and GPT 4.0 models, as can be seen from Figure 4.11, were different from each other, since both models conceptualize the problem differently. At first glance, both AI-generated diagrams contain both pools and lanes with different names, but it was necessary to understand whether these names represented the same entity from the functional side when compared to the reference solution.

For example, in the solution implemented by the GPT 3.5 model, Pool 'Traveller' and 'User Registration' are considered in the traveler pool. This is considered partially correct because both 'Traveler' and 'User Registration' represent two different entities in the solution. From a functional point of view, however, the lane called 'Trip Registration' is the same as 'User Registration', and the 'Health Screening' lane is the same as 'Employee', so

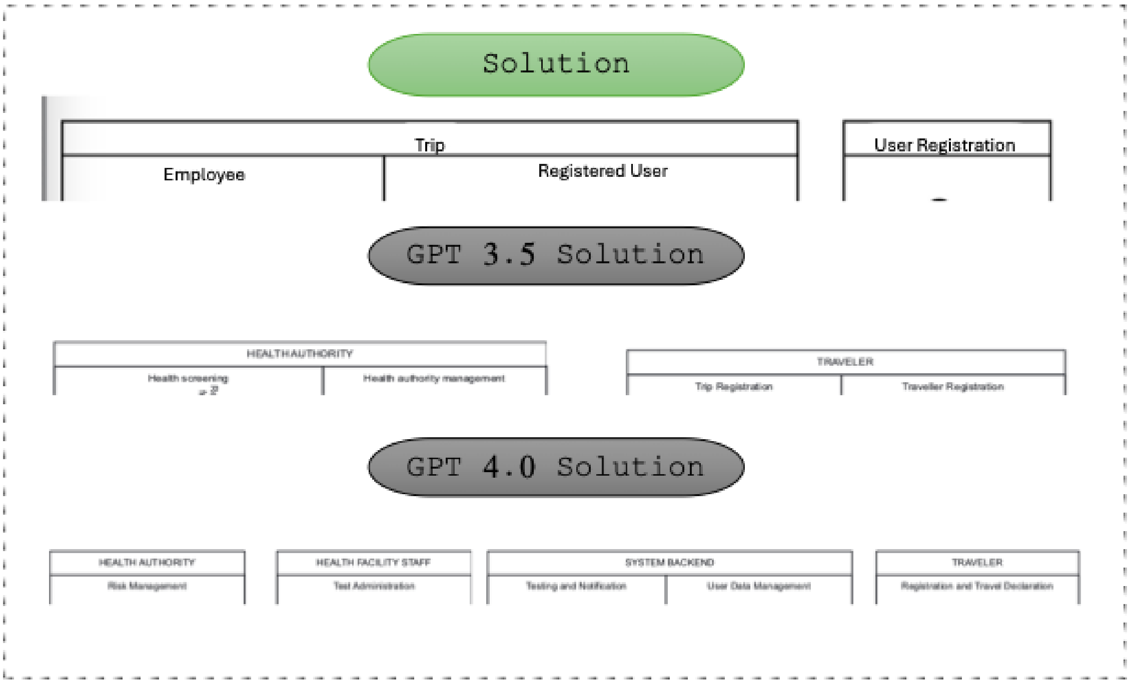

Figure 4.11. GPT 3.5 GPT 4.0 vs Solution

they can be considered correct. In the solution implemented by the GPT 4.0 model, however, the 'Trip' pool does not exist, and neither does the 'Registered User' pool. For this reason, they were assigned a 0 score.

Thus, the final points of the evaluation of pools and lanes in this exercise were as follows:

NO. CONDITION	SOLUTION	HUMAN	GPT 3.5	GPT 4.0
POOLS				
1	User Registration	1	0.5	1
2	Trip	0.5	0.5	0
LANES				
3	New user	1	1	1
4	Registered user	1	1	0
5	Employee	1	1	1

Figure 4.12. Final Scores

As can be observed, the human actor performed better than AI models, while the unexpected fact is that the GPT 3.5 model performed better than the GPT 4.0 model as it gave partially correct answers, managing to identify the main parts of the diagram. Instead, the diagram generated by GPT 4.0 had two incorrect parts, resulting in a lower overall score.

The same evaluation logic was used for the parts of the process and the order of events.

4.5.4 Statistical Methods

Another key part of this thesis is the statistical methods. After analyzing all the exam papers and evaluating them by assigning the final scores, a summary table was created with all the results. The statistical analyzes were carried out independently, one for the syntactic evaluation and one for the adequacy evaluation. This was done in two main stages:

1. Data preparation

2. Statistical analysis of the data

Before carrying out statistical analysis, the data must first be in an appropriate format.

The appropriateness evaluation dataset after the final score assignment was presented in the format below:

	HUMAN			GPT 3.5			GPT 4.0		
EXERCISE NO.	'1'=Correct	'0.5'=Partly correct	'0'=Incorrect	'1'=Correct	'0.5'=Partly correct	'0'=Incorrect	'1'=Correct	'0.5'=Partly correct	'0'=Incorrect

Figure 4.13. Structure of the appropriateness evaluation dataset for each exercise

For each actor, the respective percentage for each case, that is whether the answer was correct, partially correct, or incorrect, is reported. This dataset was not optimal, as analysing each combination of response type (correct, partially correct, or incorrect) and actor (human, GPT 3.5, and GPT 4.0) would have become too complicated and time-consuming.

The purpose of the thesis was to analyse the total performance of different actors. For this reason, to facilitate the process of studying the data, a total weighted average for each actor to obtain the total overall score was computed, using the following formula:

- Weighted Human Score =(1*Correct) +(0.5*Partly Correct))+ (0*Incorrect)
- Weighted GPT 3.5 Score =(1*Correct) + (0.5*Partly Correct) + (0*Incorrect)
- Weighted GPT 4.0 Score =(1*Correct) + (0.5*Partly Correct) + (0*Incorrect)

This way, the dataset took the following final form:

EXERCISE NO.	HUMAN	GPT 3.5	GPT 4.0
EXERCISE 1	No. of syntax errors 1	No. of syntax errors 1	No. of syntax errors 1
...	...	...	...
EXERCISE N	No. of syntax errors N	No. of syntax errors N	No. of syntax errors N

Figure 4.14. Final evaluation dataset

With the data in this format, the next step was to perform syntactic analysis of the data. To achieve this goal, two analyses were used.

1. Analysis of variance (ANOVA) and t-test for the evaluation of adequacy
2. Analysis of variance for the syntactic evaluation

The t-test is a statistical test that determines whether two samples from the same population have the same mean. In this case, since the same data variable (the exercises) was common to the three actors (human, GPT 3.5, and 4.0), the paired t-test was the appropriate test to test the hypothesis of whether being human or one of the AI models would perform better, based on accuracy scores. The t-test was run twice, a t-test comparing the performance of Human vs. GPT 3.5 and another t-test to compare the performance of Human vs. GPT 4.0.

Another statistical test used was Analysis of Variance, commonly referred to as ANOVA. ANOVA is a statistical method that can be used to compare the means of three or more samples to determine whether at least one sample is different from the others. Additionally, the analysis tests the null hypothesis that all group means are the same against the alternative hypothesis that at least one group mean is different. Depending on the variables to be studied, there are different types of analysis of variance.

In this particular case, single-factor ANOVA, which takes into account an independent variable (exercises) to evaluate whether at least the mean of one group is different from the others, was adopted. ANOVA analysis was used to compare the syntactic performance, but also the adequacy between the three actors. The main reason why this test was performed at the end (after the t-tests) was to have a more accurate statistical result, since the biggest limitation of the t-tests is that they must be performed between pairs of groups, thus increasing the risk of type I error (false positive). In this way, the possibility of obtaining inaccurate results by performing this additional test was mitigated.

4.5.5 Limitations of the Evaluation Methodology

The selected methodology for the evaluation has some limitations that can be detected, with these limitations being of two different types:

1. Potential biases
2. Constraints

The first type of limitation is related to observer biases. These biases can occur at any phase of the research, starting from the research, the choice of exercises to be analysed and lastly, the evaluation phase. Since this study was conducted primarily by one student, it raises further the potential risk of this type of bias. The rationale after this observation is that it deals with cases when a researcher's expectations or biases influence what he or she perceives and records in a study. Under these circumstances, this potential bias was mitigated by using different data sources to verify the authenticity of the results. Also necessary were the periodic checks and supervision with the supervisors of this scientific research and the continuous agreement to verify and guarantee the coherence between the results and the strategies followed in this course.

Another type of limitation are constraints that directly influence the evaluation process, with diagram complexity, the limited competence of the human solver, and the scoring system as examples. Other exercises with a different level of complexity may have yielded different results from all three actors, for example. Another issue comes from the fact that the human actor was a single student, and her knowledge on solving BPMN modeling exercises limits the results, as more or less skilled human solvers may have obtained different results. Lastly, the scoring system used to evaluate adequacy made use of three scores, but there may have been more detailed ranges of grades to adopt, considering the wide set of anomalies and errors that could be present in the diagrams.

Chapter 5

Results and Discussion

This chapter is primarily focused on presenting the raw data and findings from the study, with the Discussion section presenting an interpretation of these results.

5.1 Results

In this section, the results of the statistical analysis will be presented. As mentioned in the methodology section, two types of statistical analysis, t-test and analysis of variance, ANOVA, were used in this study.

The t-test was used to statistically analyse the accuracy scores and was used twice, once comparing the human actor with GPT 3.5, and then with GPT 4.0.

t-test results comparing humans vs. GPT 3.5 are shown in Figure 5.2:

From the results of the t-test comparing the performance of Human versus GPT 3.5 the mean average (as percentage) was found to be 0.86 for humans and 0.60 for GPT 3.5. The variance in the accuracy percentages was 0.01296 for human generated solutions and 0.01547 for GPT 3.5. The number of exercises analysed in this t-test (observations) were 19. The Pearson Correlation coefficient between the Human and GPT 3.5 accuracy percentages resulted equal to 0.2914, indicating a low positive correlation between the two sets of scores. The degrees of freedom for this t-test was 18, and is found as the number of paired observations minus 1. T-statistic value was approximately 8.0659, and it is the value used to test the hypothesis. The one-tailed p-value was 1.9003e-07, indicating a very low probability of observing such a t-statistic if the null hypothesis were true (no difference in means). The

t-Test: Paired Two Sample for Means		
	Human	*GPT 3.5*
Mean	0,863947368	0,60105263
Variance	0,012962719	0,01547939
Observations	19	19
Pearson Correlation	0,291483801	
Hypothesized Mean Difference	0	
df	18	
t Stat	8,065899536	
P(T<=t) one-tail	1,09003E-07	
t Critical one-tail	1,734063607	
P(T<=t) two-tail	2,18005E-07	
t Critical two-tail	2,10092204	

Figure 5.1. t-test Human vs GPT 3.5

two-tailed p-value was 2.18005e-07, which was also extremely low. The critical t-value for a one-tailed test at alpha level of 0.05 was 1.7340, while the critical t-value for a two-tailed test at the same alpha level was 2.1009.

Furthermore, the t-test results comparing Human vs. GPT 4.0 performance were as the following:

From the results of the t-test comparing the performance of Human versus GPT 4.0 the mean average (as percentage) was found to be 0.86 for humans and 0.59 for GPT 4.0. The variance in the accuracy percentages was 0.01296 for human generated solutions and 0.01690 for GPT 4.0. The number of exercises analysed in this t-test (observations) were 19. The Pearson Correlation coefficient between the Human and GPT 4.0 accuracy percentages resulted equal to 0.1859, indicating a low positive correlation between the two sets of scores. The degrees of freedom for this t-test was 18. T-statistic value was approximately 7.4884, and it is the value used to test the hypothesis. The one-tailed p-value was 3.0996E-07, indicating a very low probability of observing such a t-statistic if the null hypothesis were true (no difference in means). The two-tailed p-value was approximately 6.1993E-07, also suggesting a low probability.The two-tailed p-value was 2.18005e-07, which was also

t-Test: Paired Two Sample for Means		
	Human	GPT 4.0
Mean	0,86394737	0,59578947
Variance	0,01296272	0,01690629
Observations	19	19
Pearson Correlation	0,18591663	
Hypothesized Mean Difference	0	
df	18	
t Stat	7,48839717	
P(T<=t) one-tail	3,0996E-07	
t Critical one-tail	1,73406361	
P(T<=t) two-tail	6,1993E-07	
t Critical two-tail	2,10092204	

Figure 5.2. t-test Human vs GPT 3.5

extremely low. The critical t-value for a one-tailed test at alpha level of 0.05 was 1.7340, while the critical t-value for a two-tailed test at the same alpha level was 2.1009, the same as the ones registered before.

Additionally, analysis of variance was performed with the data sets obtained from accuracy scores and syntax errors.

Results of ANOVA analysis based on the data of accuracy scores are presented in Figure 5.3:

The group statistics registered respectively 19 exercises with an average accuracy of 0.8639 and variance of 0.01296 for human generated solutions, 19 exercises with an average accuracy of 0.6010 and variance of 0.01547 for GPT 3.5 generated solutions and 19 exercises with an average accuracy of 0.5958 and variance of 0.01906 for GPT 4.0 generated solutions. The sum of squares (SS) between groups with 2 degrees of freedom (df), was 0.8933 with a mean square (MS) of 0.4466. In the meanwhile, the SS within groups with 54 degrees of freedom (df) was 0.8162 that resulted in a MS of 0.0151. The F-statistic was 29.5485. The p-value associated with the F-statistic was 2.14517e-09 and the critical value of F for this test at the alpha level of 0.05 was 3.1628.

The results of the ANOVA analysis regarding the syntax error data are

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
HUMAN	19	16,415	0,863947368	0,012962719
GPT 3.5	19	11,42	0,601052632	0,015479386
GPT 4.0	19	11,32	0,595789474	0,016906287

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0,893316667	2	0,446658333	29,54845688	2,14517E-09	3,168245967
Within Groups	0,816271053	54	0,015116131			
Total	1,709587719	56				

Figure 5.3. Anova Analysis Accuracy scores

displayed in Figure 5.4:

ANOVA Analysis	Call: lmp(formula = \`Syntax Errors\` ~ Actor, data = syntax_errors, settings = FALSE) Residuals: Min 1Q Median 3Q Max -2.2632 -1.2632 -0.2632 1.1053 10.7368 Coefficients: Estimate Iter Pr(Prob) Actor1 51 0.765 Actor2 -0.22807 0.804 Residual standard error: 2.096 on 54 degrees of freedom Multiple R-Squared: 0.01155, Adjusted R-squared: -0.02506 F-statistic: 0.3156 on 2 and 54 DF, p-value: 0.7307

Figure 5.4. ANOVA analysis syntax errors.

The minimum, 1st quartile, median, 3rd quartile, and maximum values provide a five-number summary of the residuals which are found as the differences between observed and predicted values. The range of residuals indicate the spread of the errors around the predicted values. The estimate for 'GPT 3.5 vs Human' is 0.765, which suggests that the mean value of 'Syntax Errors' for GPT 3.5 is 0.7665 units higher than that of the human. Instead, the estimate for 'GPT 4.0 vs Human' is -0.22807, suggesting that the mean of GPT 4.0 is 0.22807 units lower than that of human. The residual standard error is 2.096 on 54 degrees of freedom, giving a measure of the spread of the residuals. The R-squared value is 0.0155, and it suggests that approximately only 1.55 % of the variability in syntax errors is explained by the actor factor. The adjusted R-squared value is 0.1250. The F-statistic is 0.3156 on 2 and 54 degrees of freedom while the p-value associated with it is 0.7307, and it is much higher than the alpha level of 0.05.

5.2 Discussion

After analyzing the data, it is clear that the performance of the three actresses included in this study is different. At the beginning of this study, a central null hypothesis was presented:

'The performance of the three actors(human, GPT 3.5 and GPT 4.0) in the study is the same, which implies that there is no difference in performance.'

After the statistical analysis, it was clear that humans performed better than the GPT models regarding accuracy scores. The surprising part is that, from the ANOVA analysis for syntax errors, the GPT 3.5 model had a better mean score than the human one. However, the null hypothesis was rejected as the p-value was lower than the conventional alpha of 0.05.

The nature of this work was experimental, meaning that all the initial expectations were that the human model and the GPT model would have the same performance. Conversely, the results showed the opposite; at the moment, the GPT models cannot generate the most accurate diagram of those generated by human beings as students or experts in the sector. With the rapid development of artificial intelligence, the AI generation of BPMN diagrams remains an open research question. Unfortunately, no proof or other scientific articles have been tried to address this aspect, and only research describes the process of adapting the technology to this process.

This study itself presented two different aspects, being that if, on the one hand, it can be used as scientific evidence in the field of generating diagrams, on the other hand, there are limitations. One of the strong points is the data set and its diversity. Since the problem descriptions used were exam papers used to test the student's knowledge, they presented a complex nature in terms of thinking and logic. This made it possible to test the AI models' thinking capacity since each problem description offered to the model was different and had different degrees of difficulty. This could explain the low accuracy scores. Another strong point of the methodology was its iterative nature, subject to continuous improvements.

On the other hand, the study design and methodology had various limitations. The first and most important limitation is that only one student solved the human-generated diagrams. This is a critical limitation since human performance is mainly based on a person's knowledge, which can often be guided by subjective reasoning. Secondly, the other limitation is that the AI models could not generate the graphic representation of the diagram directly but could only reproduce the logic part and the graphic construction was done by the student conducting this dissertation work. The open question would be to define the level of the margin of human error in the graphic contextualization of the diagrams and their evaluation.

Future research can be channelled in several directions to address the abovementioned limitations. First, from the technology suitability side, a preliminary process of model training in specific knowledge of BPMN notations and knowledge regarding the generation of the graphic part can be considered. This would be challenging since the work would have to be directed on three main fronts:

- training the model for BPMN notations
- training the model to generate XML files that contain the graphic side
- training the model to reason logically, accurately and compliant with the above requirements

In addition to the variety of problem descriptions or scenarios, the development of this study with more students can be considered in an attempt to reduce potential bias.

This study's feature and contribution is that it tries to provide an empirical test of a research question that currently lacks direct scientific evidence. BPMN diagrams are an essential part of process modelling, and this study focused on the possibility that AI models could serve as an aid to the domain experts who construct these diagrams; why not their automatic creation, helping to automate this process in an AI-driven way.

BPMN diagrams are part of the business analysis of every business. However, AI automation in creating these diagrams would also bring some ethical concerns related to the use of the technology. The first and most crucial concern is data security and privacy. Business processes are subject to a high level of privacy since, often, the way a company gains a competitive advantage over another is influenced by the way it manages and optimizes processes. Giving AI access to analyzing processes and then generating diagrams would have a significant impact if there was not maximum security in data protection.

On the other hand, other concerns may be related to bias, fairness, explainability, and transparency in these models. The fact that AI models are trained from internet-based data means that there is a possibility that they carry biases in this data. On the other hand, AI models are built from a set of algorithms that are not accessible by the user; in these cases, there is no transparency in the process of generating the answer. These two concerns are directly related to whether businesses should trust the results and diagrams generated.

This research work brought out a difficulty which was not foreseen in the generation of diagrams. The biggest challenge was that the AI models could not generate the graphical representation of the diagrams, and the generated XML files were not coherent and consistent. This required the adaptation of a new methodology, highlighting this limit of AI models.

Chapter 6

Evaluations

6.1 Evaluation of AI-generated diagrams

In this section, the results and quality of the diagrams constructed by the GPT 3.5 and GPT 4.0 models, as well as the differences between them, will be comprehensively analysed and interpreted. Two interpretations will be made taking into account three different points of view: quality of the diagrams, adherence to standards and usability and interpretability.

To begin with, regarding the quality of the diagrams created by AI models, the biggest problem and challenge faced is the fact that the models were not always able to capture the key components and relationships described in the problem statements. What was observed was that when the problems had a complex structure or contained a lot of information within it, the models were not accurate in generating an optimal answer, but it was necessary to provide additional information to generate an answer. Naturally this activity was a time consuming task. As demonstrated by the empirical appropriateness evaluation data, in relation to problem solving, the AI models were respectively able to capture approximately 60% of the main aspects of the requested problem.

Qualitatively, the diagrams generated from the GPT 3.5 and GPT 4.0 models had some substantial differences between them. If it was to judge the complexity of the diagrams, certainly the diagrams generated by GPT 3.5 were simpler than those generated by GPT 4.0. This is also expected since GPT 4.0 is trained on a larger dataset and its specific domain knowledge may be greater. But what was not expected was that the diagrams generated by GPT 3.5 would be more syntactically correct than those generated by GPT 4.0. In the final results, the GPT 3.5 model had a total of 33 errors across all exercises compared to the 43 errors in GPT 4.0, a significant difference.

An important point to underline is that, when using the term diagrams generated by artificial intelligence models, the actual meaning is a textual representation that contains the logical sequence of activities. The actual diagram was constructed by humans as the AI models were unable to generate an adequate graphical representation.

From a logical point of view, it can be said that both models succeeded in producing a logical sequence. GPT 3.5 was successful in maintaining the logical sequence between tasks in most cases, although these solutions had missing parts and the model was unable to include all aspects of the diagram. For example, in most cases, the GPT 3.5 model rarely (almost never) identified more than one pool of a generated solution and even message flows. In most cases this was a limitation of the model. Another limitation was the fact that the model was often able to identify multiple lanes, but in the final solution it was not able to identify any part of the process belonging to a specific lane or pool, as can be seen from Figure 6.1. This limitation constituted a fairly frequent error too, as the model apparently was unable to maintain consistency with the generated solution.

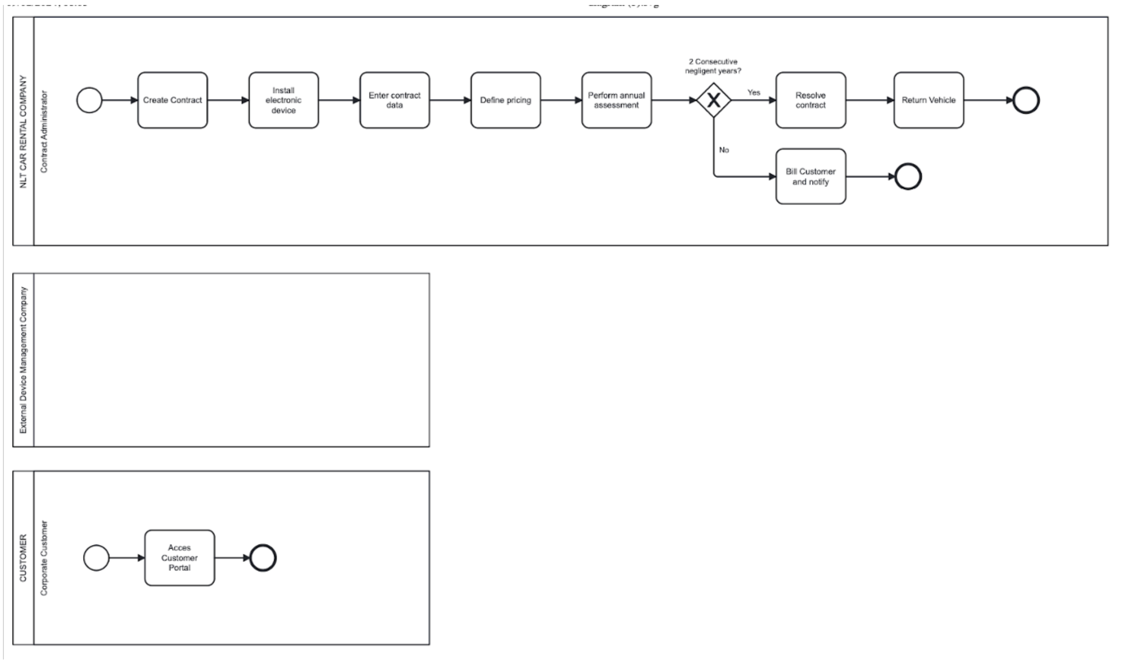

Figure 6.1. Diagram generated from model GPT 3.5

In contrast to the 3.5 model, the 4.0 model had difficulty maintaining sequential logical coherence between tasks, processes, and flows. The 4.0 model was able to identify message flows and pools in most cases, but sometimes did not provide the source and destination activity of this flow. The interesting thing was that when the user queried it again, perhaps asking for additional information, instead of integrating the missing part, GPT 4.0 often lost the logical flow with the previously generated response. Again, this operation constituted a time-consuming task as well as a limitation of the 4.0 model. Figure 6.2 presents an example of this situation, with a message collection event but no corresponding message generating event present in the diagram.

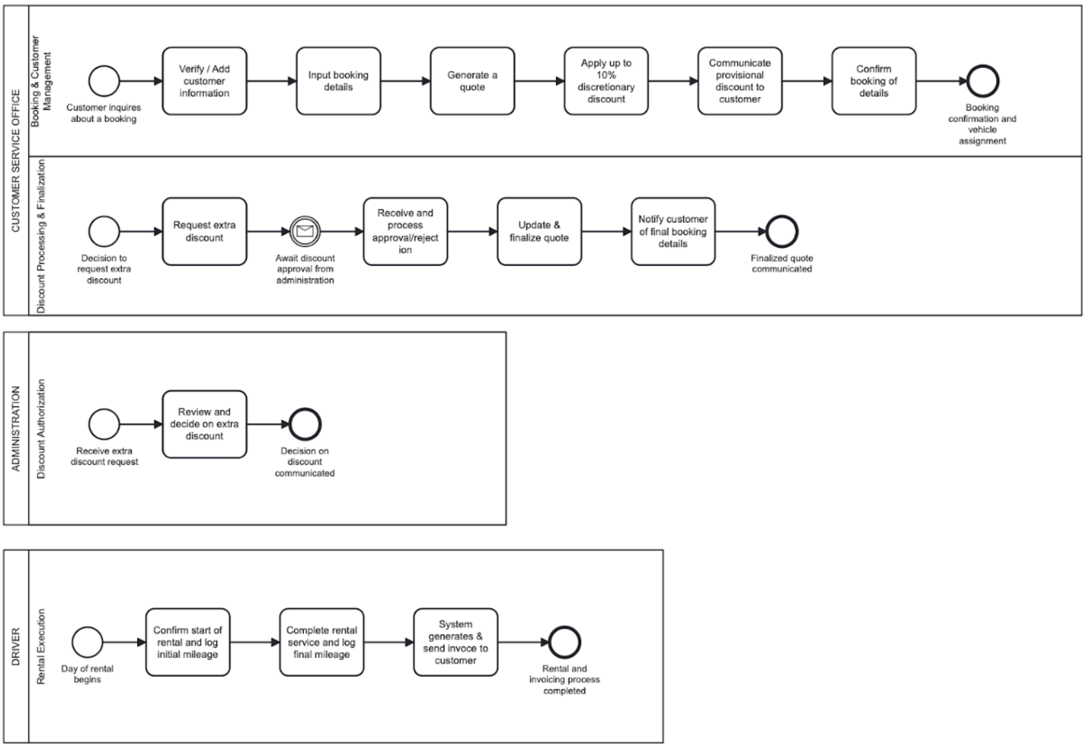

Figure 6.2. Diagram generated from model GPT 4.0

Regarding the concept of adherence to standards, both models adapt to the BPMN 2.0 standards to some extent. An important role in this part is also played by the student who constructed the diagrams, since apart from the final suggestion, sometimes the models needed additional instructions and guidance to generate an optimal response compliant with the standards.

Some of the most frequent logical errors made by AI models when generating diagrams are shown in Figure 6.1:

Table 6.1. Most common logical errors found in the GPT-made diagrams

GPT 3.5	GPT 4.0
No message flows	No explicit message flows
No timer events	Incomplete gateways
Incomplete gateways	Between the same pool, non continuity in the tasks
Identification of pools and lanes with no activity in it	

From an interpretability perspective, the AI-generated diagrams were easily readable, understandable, and interpretable by someone who has standard knowledge of these diagrams. Normally both kinds of diagrams were not complete in terms of accuracy, so they may give a partial understanding of the information system in question, but in general they were readable. It is important to mention that these diagrams were used for comparative reasons only, but were not intended for a specific audience. Under these conditions no real study has been done on the quality of these diagrams from the semantic and pragmatic side.

6.2 Comparative Evaluation

In this section it will be discussed whether, in general, the diagrams generated by the AI models were more qualitative than those human-generated. As was presented in the previous chapter, in terms of quality, artificial intelligence models are currently not capable of generating more accurate diagrams than human ones.

As can be seen from Figure 6.3, the human-generated diagrams had higher performance scores than those AI-generated in almost all aspects, starting from the identification of pools and lanes, identification of process parts and sequential logical order of events.

On the other hand, what was not expected was that GPT 3.5 was syntactically more accurate than the human actor. Figure 6.4 clearly shows that GPT 3.5 performed better syntactically with only 30% errors followed by human with 32% and GPT 4.0 with 38%.

Another interesting aspect to observe and discuss is the coherence and variability between the diagrams generated by the three actors. For the AI models, the same starting prompt was used precisely to distinguish how similar the diagrams were to each other when starting from the same starting

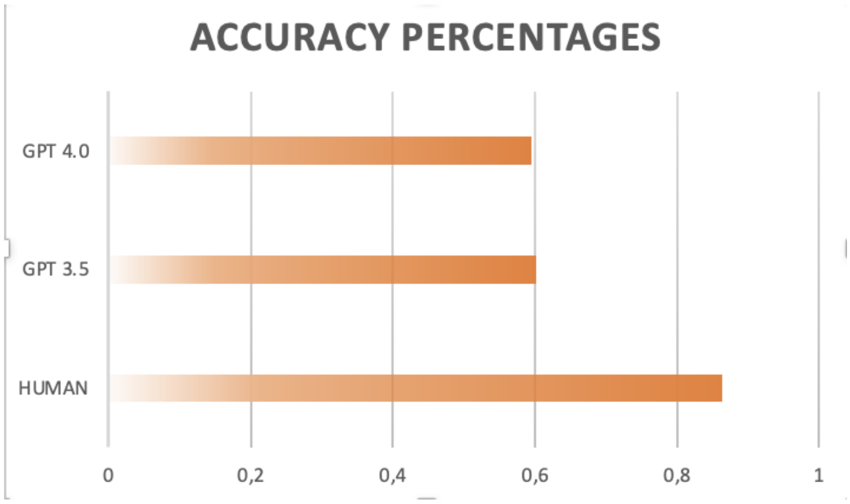

Figure 6.3. Accuracy percentages between actors

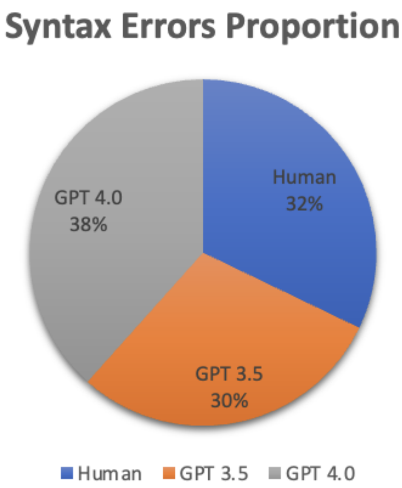

Figure 6.4. Syntax errors percentages

prompt. The results, both factual data and observations, showed that the variability between the diagrams generated by GPT 3.5 and GPT 4.0 was moderately high.

For the same problem statement, sometimes the response generated by GPT 4.0 was completely different from that of GPT 3.5. The metrics used for their evaluation were comparative with respect to the solution of the problem, and this was a bit limiting in the overall evaluation of the diagram, given that many times GPT 4.0 generated a diagram close to that of the solution and perhaps was non-compliant to the solution, but this does not mean that it is fundamentally wrong. This would require further investigation and study.

On the other hand, the variability between the diagrams generated by artificial intelligence models also has to do with the nature of artificial intelligence itself. This thesis was conducted over a period of approximately 6 months and diagram solutions are not necessarily generated on the same day. AI models are constantly being improved, which is why the models have not always been able to generate consistent responses with each other. Furthermore, the nature of AI is also random since the models on which it is based to generate outputs also have randomness, for this reason it was clearly observable that AI generated different responses at different times. This feature of artificial models can be considered an advantage and a disadvantage. Advantage because from this quality AI can introduce new forms of creativity and innovation, on the other hand it is a disadvantage since this creates variability in the answers given.

The variability in human-generated diagrams has to do with the amount of exercises solved over time. The more experience is gained, the more specialised the solver will become in that task and therefore subsequent diagrams will be more accurate. In terms of creativity and innovation, it is quite certain that two different people have a different perception and interpret a diagram differently, and perhaps they can give a different solution and interpretation. After this thesis, it cannot be said whether there is a noticeable difference between the diagrams created by humans and those of artificial intelligence models.

Based on the results of the evaluation, statistically, it cannot be said that AI is currently capable of replacing humans in the generation of BPMN diagrams. However, considering the results obtained from the study in a broader context, the data is promising and in the future, with the expected increase in the capacity of artificial intelligence, it would be interesting to see whether artificial intelligence can occupy an important place in the field of software and process modelling.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

The results obtained provide an answer to the initial hypothesis presented at the beginning of this research work but also serve as an empirical test, providing statistical proof of a still open scientific question.

As mentioned above, the results obtained from the statistical analyzes (t-test and ANOVA) rejected the null hypothesis that AI and human models have the same performance in generating diagrams. The empirical analysis and the value of the p-value, which was lower than the conventional alpha, shows that there is no statistical evidence to verify the initial hypothesis. This conclusion was unexpected, as initial expectations were that AI models would generally perform better than they actually did.

In addition, the results obtained from this study shed light on a field which is still new and little studied by experts in the field, since there is still no genuine research that provides empirical evidence for this aspect. On the other hand, they highlight a limit of AI models, that of not generating a BPMN diagram in a complete and coherent XML file, especially when the problem is long and complex. This means that further studies are still needed to understand the capacity of AI models from this aspect.

However, it is important to highlight the limitations that may impact the interpretation and generalization of the results. These limitations can be technical or methodological. The first type concerns the technical ability of the AI models to give the desired result, while the second type concerns the methodology chosen to conduct this study, which may be subject to interpretation.

7.2 Future Work

In conclusion, this research on the role of AI in the generation of BPMN diagrams clearly provides an answer to an open research question and sheds light on new questions that need further exploration.

One could be improving the methodology used, for example, training AI models or using a different AI model. Since this study mainly focused on evaluating and comparing syntactic errors, an exciting aspect would be if the generated diagrams were semantically and pragmatically correct. To answer this question, broader and more specialized human resources expertise would be needed. On the other hand, it was noticed that during the generation of the logical workflow, the performance of the artificial intelligence was sometimes inconsistent. Having insight into how model architectures and algorithms behind AI models are built would help mitigate the risk of variability and inconsistency in the responses generated by these models when given different problems.

The technology of artificial intelligence in the generation of diagrams can create new opportunities for many new technological developments. For example, the technology industry can benefit greatly from these developments since they can streamline the development process, while sectors focused on improving processes and logistics can use them to improve processes. The creation of applications for the academic field that create various diagrams and schemes for summarizing the text could also be a further development.

Finally, these findings can also be used in real-life contexts. For example, the IT industry would benefit the most from AI-generated diagrams. This technological advance would simplify the modelling process, reduce errors, save time, and allow the automation and documentation of this process.

Bibliography

[1] "model, v. meanings, etymology and more | Oxford English Dictionary." *Oed.com*, 2024. [online] https://doi.org/10.1093//OED//6214240665.

[2] Gomaa, H. A Software Modeling Odyssey: Designing Evolutionary Architecture-Centric Real-Time Systems and Product Lines. *Lecture Notes in Computer Science*, pp.1–15, 2006. https://doi.org/10.1007/11880240_1.

[3] Gomaa, H. *Software modeling and design : UML, use cases, patterns, and software architectures*. Vancouver, B.C.: Langara College, 2018.

[4] Rumbaugh, J. Unified Modeling Language (UML). pp.1249–1257, 2010. https://doi.org/10.1081/e-ese-120044214.

[5] Choi, M.-Y., Lim, J., and Joo Kyung-Soo. Developing an XML Schema generator based on UML class. 2005. https://doi.org/10.1109/icmlc.2005.1527334.

[6] Dumas, M., La Rosa, M., Mendling, J. and Reijers, H.A., 2013. *Fundamentals of Business Process Management*. Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-33143-5.

[7] D'Ippolito, B., 2014. The importance of design for firms' competitiveness: A review of the literature. *Technovation*, [online] 34(11), pp.716–730. https://doi.org/10.1016/j.technovation.2014.01.007.

[8] von Rosing, M., White, S., Cummins, F., and de Man, H. Business Process Model and Notation—BPMN. In *The Complete Business Process Handbook*, pp.433–457, 2015. https://doi.org/10.1016/b978-0-12-799959-3.00021-5.

[9] Bedwell, K., Garaccione, G., Coppola, R., Ardito, L., and Morisio, M. BIPMIN: A Gamified Framework for Process Modeling Education. *Information*, 14(1), p.3, 2022. https://doi.org/10.3390/info14010003.

[10] Fan, L., Li, L., Ma, Z., Lee, S., Yu, H. and Hemphill,

L., 2023. A Bibliometric Review of Large Language Models Research from 2017 to 2023. *arXiv* (Cornell University). https://doi.org/10.48550/arxiv.2304.02020.

[11] Douglas, M.R., 2023. Large Language Models. https://doi.org/10.48550/arxiv.2307.05782.

[12] Uszkoreit, J., 2017. Transformer: A Novel Neural Network Architecture for Language Understanding. [online] Google AI Blog. Available at: https://ai.googleblog.com/2017/08/transformer-novel-neural-network.html.

[13] Song, X. and Xiong, T. A Survey of Published Literature on Conversational Artificial Intelligence. [online] IEEE Xplore, 2021. https://doi.org/10.1109/ICIM52229.2021.9417135.

[14] Darko, A., Chan, A.P.C., Adabre, M.A., Edwards, D.J., Hosseini, M.R., and Ameyaw, E.E. Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities. *Automation in Construction*, 112, p.103081, 2020. https://doi.org/10.1016/j.autcon.2020.103081.

[15] Khanzode, K.C.A. and Sarode, R.D., 2020. Advantages and disadvantages of artificial intelligence and machine learning: A literature review. *International Journal of Library & Information Science (IJLIS)*, 9(1), pp.3.

[16] Blanchard, A. and Taddeo, M., 2023. The Ethics of Artificial Intelligence for Intelligence Analysis: a Review of the Key Challenges with Recommendations. *Digital Society*, [online] 2(1). https://doi.org/10.1007/s44206-023-00036-4.

[17] Stahl, B.C., Schroeder, D. and Rodrigues, R., 2022. The Ethics of Artificial Intelligence: A Conclusion. In *Ethics of Artificial Intelligence*, pp.107–111. https://doi.org/10.1007/978-3-031-17040-9_9.

[18] Zawacki-Richter, O., Marín, V.I., Bond, M., and Gouverneur, F. Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, [online] 16(1). https://doi.org/10.1186/s41239-019-0171-0, 2019.

[19] Zhai, X., Chu, X., Chai, C.S., Jong, M.S.Y., Istenic, A., Spector, M., Liu, J.-B., Yuan, J., and Li, Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. *Complexity*, [online] 2021(8812542), pp.1–18. https://doi.org/10.1155/2021/8812542, 2021.

[20] Ma, X. and Jiang, C. On the Ethical Risks of Artificial Intelligence Applications in Education and Its Avoidance Strategies. *Journal of Education Humanities and Social Sciences*, 14, pp.354–359, 2023.

https://doi.org/10.54097/ehss.v14i.8868.

[21] Kamalov, F., Calonge, D.S., and Gurrib, I. New Era of Artificial Intelligence in Education: Towards a Sustainable Multifaceted Revolution. *Sustainability*, [online] 15(16), p.12451, 2023. https://doi.org/10.3390/su151612451.

[22] Zhai, X., Chu, X., Chai, C.S., Jong, M.S., Starčič, A.I., Spector, M., Liu, J., Jing, Y., & Li, Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020, 2021, April 20. https://doi.org/10.1155/2021/8812542.

[23] Hwang, G., & Tu, Y. Roles and Research Trends of Artificial Intelligence in Mathematics Education: A Bibliometric Mapping Analysis and Systematic Review. *Mathematics*, 9(6), 584-584, 2021, March 10. https://doi.org/10.3390/math9060584.

[24] Li, B., Hou, B., Yu, W., Lu, X., and Yang, C. Applications of artificial intelligence in intelligent manufacturing: a review. *Frontiers of Information Technology & Electronic Engineering*, [online] 18(1), pp.86–96, 2017. https://doi.org/10.1631/fitee.1601885.

[25] Kamran, S.S., Haleem, A., Bahl, S., Javaid, M., Prakash, C., and Budhhi, D. Artificial intelligence and advanced materials in automotive industry: Potential applications and perspectives. *Materials Today: Proceedings*, [online] 62, 2022. https://doi.org/10.1016/j.matpr.2022.04.727.

[26] Hofmann, M., Neukart, Florian, and Bäck, T. Artificial Intelligence and Data Science in the Automotive Industry. arXiv (Cornell University), 2017. https://doi.org/10.48550/arxiv.1709.01989.

[27] Neuroth, M., MacConnell, P., Stronach, F., and Vamplew, P. Improved modelling and control of oil and gas transport facility operations using artificial intelligence. *Knowledge-Based Systems*, 13(2-3), pp.81–92, 2000. https://doi.org/10.1016/s0950-7051(00)00049-6.

[28] Alvarado, M. Autonomous agents and computational intelligence: the future of AI application for petroleum industry. *Expert Systems with Applications*, 26(1), pp.3–8, 2004. https://doi.org/10.1016/s0957-4174(03)00103-9.

[29] Koroteev, D. and Tekic, Z. Artificial intelligence in oil and gas upstream: Trends, challenges, and scenarios for the future. *Energy and AI*, 3, p.100041, 2021. https://doi.org/10.1016/j.egyai.2020.100041.

[30] Laska, (2023). Artificial intelligence in the chemical industry – risks and opportunities. *Scientific Papers of Silesian University of Technology Organization and Management Series*, [online] 2023(172).

https://doi.org/10.29119/1641-3466.2023.172.25.

[31] Liao, M., Lan, K., and Yao, Y. Sustainability implications of artificial intelligence in the chemical industry: A conceptual framework. *Journal of Industrial Ecology*. https://doi.org/10.1111/jiec.13214, 2021.

[32] Luckow, A., Kennedy, K., Ziolkowski, M., Djerekarov, E., Cook, M., Duffy, E., Schleiss, M., Vorster, B., Weill, E., Kulshrestha, A., and Smith, M.C. Artificial Intelligence and Deep Learning Applications for Automotive Manufacturing. [online] *IEEE Xplore*, 2018. https://doi.org/10.1109/BigData.2018.8622357.

[33] Hofmann, M., Neukart, Florian, and Bäck, T. Artificial Intelligence and Data Science in the Automotive Industry. arXiv (Cornell University), 2017. https://doi.org/10.48550/arxiv.1709.01989.

[34] Dubey, V. and Walimbe, R.S. Fintech 2022 Trends: The era of F-CUBE 'Fast and Furious Fintech'. *Technium: Romanian Journal of Applied Sciences and Technology*, 3(9), pp.44–50, 2021. https://doi.org/10.47577/technium.v3i9.4976.

[35] Thomas, R. (2021). Introduction: Building the AI bank of the future. [online] *McKinsey & Company*. Available at: https://www.mckinsey.com/industries/financial-services/our-insights/introduction-building-the-ai-bank-of-the-future [Accessed 24 Mar. 2024].

[36] Gonçalves, Ana Rita, Meira, Amanda Breda, Shuqair, Saleh, and Pinto, Diego Costa. Artificial intelligence (AI) in FinTech decisions: the role of congruity and rejection sensitivity. 2023. https://doi.org/10.1108/ijbm-07-2022-0295.

[37] Yeh, H.-C., Yu, M.-C., Liu, C.-H., and Huang, C.-I. Robo-advisor based on unified theory of acceptance and use of technology. *Asia Pacific Journal of Marketing and Logistics*. https://doi.org/10.1108/apjml-07-2021-0493, 2022.

[38] Hu, X. and Wang, K. Bank Financial Innovation and Computer Information Security Management Based on Artificial Intelligence. [online] *IEEE Xplore*, 2020. https://doi.org/10.1109/MLBDBI51377.2020.00120.

[39] Hu, Z. Research on Fintech Methods Based on Artificial Intelligence. *Journal of Physics: Conference Series*, 1684, p.012034, 2020. https://doi.org/10.1088/1742-6596/1684/1/012034.

[40] Xie, M. Development of Artificial Intelligence and Effects on Financial System. *Journal of Physics: Conference Series*, 1187(3), p.032084, 2019. https://doi.org/10.1088/1742-6596/1187/3/032084.

[41] Ashta, A. and Herrmann, H. Artificial intelligence and fintech: An overview of opportunities and risks for banking, investments, and microfinance. *Strategic Change*, 30(3), pp.211–222, 2021. https://doi.org/10.1002/jsc.2404.

[42] Azzi, S., Gagnon, S., Ramirez, A., and Richards, G. Healthcare Applications of Artificial Intelligence and Analytics: A Review and Proposed Framework. *Applied Sciences*, 10(18), p.6553, 2020. https://doi.org/10.3390/app10186553.

[43] Fogel, A.L. and Kvedar, J.C. Artificial intelligence powers digital medicine. *npj Digital Medicine*, 1(1), 2018. https://doi.org/10.1038/s41746-017-0012-2.

[44] Choudhury, A. and Asan, O. Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review. *JMIR Medical Informatics*, [online] 8(7), p.e18599, 2020. https://doi.org/10.2196/18599.

[45] Bajwa, J., Munir, U., Nori, A., and Williams, B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*, [online] 8(2), pp.e188–e194, 2021. https://doi.org/10.7861/fhj.2021-0095.

[46] Lee, D. and Yoon, S.N. Application of Artificial Intelligence-Based Technologies in the Healthcare Industry: Opportunities and Challenges. *International Journal of Environmental Research and Public Health*, 18(1), p.271, 2021.

[47] Safieddine, F. and Nakhoul, I. Generic Business Process Model for SMEs in M-Commerce Based on Talabat's Case Study. *Lecture Notes in Computer Science*, pp.264–278, 2018. https://doi.org/10.1007/978-3-030-02131-3_24.

[48] Stravinskiene, I. and Serafinas, D. The Link between Business Process Management and Quality Management. *Journal of Risk and Financial Management*, 13(10), p.225, 2020. https://doi.org/10.3390/jrfm13100225.

[49] Segatto, M., Inês Dallavalle de Pádua, S. and Pinheiro Martinelli, D. Business process management: a systemic approach? *Business Process Management Journal*, 19(4), pp.698–714, 2013. https://doi.org/10.1108/bpmj-jun-2012-0064.

[50] Wohed, P., van der Aalst, W.M.P., Dumas, M., ter Hofstede, A.H.M. and Russell, N. On the Suitability of BPMN for Business Process Modelling. *Lecture Notes in Computer Science*, pp.161–176, 2006. https://doi.org/10.1007/11841760_12.

[51] Mungoli, N. and Charlotte, U. Exploring the Synergy of Prompt Engineering and Reinforcement Learning for Enhanced Control and Responsiveness in Chat GPT. *J Electrical Electron Eng*, [online] 2(3), pp.201–205, 2023a. Available at: https://www.opastpublishers.com/open-access-articles/exploring-the-synerg
[Accessed 1 Mar. 2024].

[52] Velásquez, J.D., Franco, C. and Cadavid, L. Prompt Engineering: a methodology for optimizing interactions with AI-Language Models in the field of engineering. *Dyna-colombia*, 90(230), pp.9–17, 2023a. https://doi.org/10.15446/dyna.v90n230.111700.