

**International Journal of Intelligent Engineering Informatics**

ISSN online: 1758-8723 - ISSN print: 1758-8715

<https://www.inderscience.com/ijiei>

---

**BP-DSS3: a business processes decision-support chatbot through the language model GPT-3.5-turbo**

Chaima Afifi, Ali Khebizi, Khaled Halimi

**DOI:** [10.1504/IJIEI.2025.10071287](https://doi.org/10.1504/IJIEI.2025.10071287)

**Article History:**

|                   |                 |
|-------------------|-----------------|
| Received:         | 09 May 2024     |
| Last revised:     | 16 October 2024 |
| Accepted:         | 16 October 2024 |
| Published online: | 21 May 2025     |

---

## **BP-DSS3: a business processes decision-support chatbot through the language model GPT-3.5-turbo**

---

**Chaima Afifi\***

Laboratory of LABSTIC,  
University of 8 May 1945,  
Guelma, Algeria  
Email: afifi.chaima@univ-guelma.dz

\*Corresponding author

**Ali Khebizi and Khaled Halimi**

Department of Computer Science,  
University of 8 May 1945,  
Guelma, Algeria  
Email: khebizi.ali@univ-guelma.dz  
Email: halimi.khaled@univ-guelma.dz

**Abstract:** Nowadays, deploying automated decision-making tools is a crucial requirement for conducting successful business activities and efficient transactions. Hence, leveraging emerging artificial intelligence (AI) techniques, conjugated with natural language processing (NLP) will strengthen the decision support systems (DSS) that are imperative for elaborating strategic planning. In this perspective, the deployment of the emerging chatbots' technology in the area of business processes (BP) offers large capabilities for decision-making activities and yields several benefits. In fact, the ability of these conversational tools to handle and interpret textual data reflecting BP executions significantly enriches the DSS ecosystem by extending its performances with a valuable layer that improves decision-making tasks. In this work, the potential of NLP techniques to build a generative pre-trained transformer (GPT)-based chatbot is exploited. The developed conversational agent, named BP-DSS3, fine-tunes the GPT-3.5 Turbo language model and constitutes an efficient tool that supports and facilitates firm decision-making activities.

**Keywords:** business processes; generative pre-trained transformer; GPT; fine-tuning; decision-making; language model; natural language processing; NLP.

**Reference** to this paper should be made as follows: Afifi, C., Khebizi, A. and Halimi, K. (2025) 'BP-DSS3: a business processes decision-support chatbot through the language model GPT-3.5-turbo', *Int. J. Intelligent Engineering Informatics*, Vol. 13, No. 5, pp.1–31.

**Biographical notes:** Chaima Afifi is a student at the Computer Science Institute, 8 Mai 1945 University, Guelma, Algeria. Her affiliation with the LabStic Laboratory is with the ‘web technologies’ research group. She is a researcher in artificial intelligence and is working for a doctorate in business process.

Ali Khebizi is currently working at the Computer Science Institute, 8 Mai 1945 University, Guelma, Algeria. He holds PhD at the Annaba University in 2017. He is a member of the research group ‘web technologies’ in the LabStic Laboratory. He does his research in advanced information systems and integration, web services, business processes modeling and verification, distributed computing and databases. His most recent publication is ‘A framework for evaluating migration strategies of evolving business processes’.

Khaled Halimi is currently working as a teacher of Computing at the Computer Science Department of Guelma University (Algeria) and he is a researcher at the LabSTIC laboratory. He holds his PhD at Annaba University. His research fields are collaborative learning, social semantic web and social networks.

## 1 Introduction

Today’s business activities are characterised by perpetual changes that occur in firms’ environment at an increasing rate. Further, modern organisations evolve in a more complex and a competitive context requiring a paradigm shift for performing their daily BPs (Khebizi et al., 2017). In fact, the complex interaction of multiple factors, ranging from globalisation phenomena and market dynamics to internal resources management issues, requires the deployment of sophisticated solutions, and efficient approaches and techniques. To this end, integrating automated tools for supporting decision-making related to the companies’ activities rises as the ultimate resort (Bharadiya, 2023).

In this perspective, the latest industrial revolution Industry 5.0 emphasises human-machine collaboration and incorporates the human aspect into industrial processes and socio-economic activities. Thus, it promotes customisation and adaptability while leveraging technologies, such as the internet of things (IoT), and AI techniques for increased efficiency and productivity. In particular, the integration of NLP techniques (Shamshiri et al., 2024) into DSS systems appears to be a promising and groundbreaking way.

Alongside the recent technological developments and advances, conversational tools have recently invaded different areas of human activities, both in the research and in the economic spheres. In particular, chatbots tend to gain popularity very quickly and their usage is becoming more widespread. The observed success of such software is induced, particularly, by the qualitative leap achieved by the GPT technology. Especially, the democratised and ubiquitous usage of the ChatGPT tool is becoming a spectacular phenomena. ChatGPT is a cutting-edge conversational AI model developed by OpenAI and refined using the powerful GPT-3.5-turbo architecture (Jusman et al., 2023). The symbiotic partnership between ChatGPT and GPT-3.5-turbo represents current trends in NLP. This junction has the strong potential to achieve previously unheard-of levels of

success for numerous application domains, particularly, in the business processes (BPs) management area.

The objective of this work consists of investigating the adoption of innovative language models to improve decision-making in the context of BPs. The NLP techniques tailored to BPs' DSS are exploited, and a conversational chatbot whose capabilities are fine-tuned with domain specific data is developed. Thus, instead of dealing with complex graphical representations and formal models when managing and analysing BPs activities, resorting to the use of natural languages can significantly improve BPs understanding and mastering. Hence, relying on NLP allows facilitating the selection of the appropriate decision to be adopted at each progression stage of the BP execution. In such scenarios, BP managers can assess the extent to which particular BPs structures and functionalities can impact strategic decision making. In addition, by consolidating the BP ecosystem with effective decision support tools, the communication and collaboration activities between the different involved actors are strengthened.

To achieve the previous goals, this study focuses on the capabilities of NLP and exploits their potential to provide decision-makers with a more comprehensive and effective decision-support framework. Such directions are motivated by the fact that traditional decision support tools are often limited to tasks related to the management of complex BP, accurate business parameters, and the management of critical and interactive circumstances of systems (van Dun et al., 2023). However, real-time requirements, conjugated with the technical complexity of handling the vast amount of heterogeneous and distributed data, as well as the need for BPs integration and collaboration involve more innovative solutions and tools that to be deployed by BPs managers and used by decision-makers.

Leveraging the ability of language models to support and understand natural languages and their structures has numerous advantages. Thus, they constitute a potential way able to meet the previous challenges and overcoming the inherent limitations of decision-making. In fact, language models are a promising way to be adopted for bridging the previous gap by providing adequate and sufficient answers related to decision-making concerns and challenges in the context of BPs.

By taking advantages of the power of GPT-3.5-turbo, this work aims to enhance decision-making tasks. It allows facilitating natural language conversations, automating routine management procedures, and offers intelligent guidance based on understanding and analysing the vast stored amounts of knowledge reflecting BPs execution. The ultimate goal of the developed innovative solution consists to support decision-makers with advanced AI capabilities, while improving the efficiency, effectiveness, and adaptability of BP decision-making in complex and dynamic environments. In this perspective, the main contributions of this work are the following.

- 1 A thorough investigation of the groundbreaking potential of fine-tuned chatbots for customised and specific tasks is conducted.
- 2 The development of a conversational chatbot BP-DSS3 tailored to decision-making in the context of BPs.
- 3 The exploration of a wide range of comprehensive datasets representing various complex decision points encountered in different business scenarios for realising the fine-tuning procedure of the conceived chatbot.

- 4 Finally, by analysing the introduced training data, the developed conversational tool BP-DSS3 is taught to recognise and intelligently react to complex queries related to decision-making during the BP life-cycle.

The remainder structure of the paper is organised as follows. Section 2 exposes the related works and the motivations having led to the current research work. The problem specification and the description of the proposed framework are depicted in Section 3, while Section 4 exposes and discusses the experimental results. Finally, Section 5 concludes the paper and outlines our future research perspectives.

## 2 Related work

In recent years, the AI techniques have invaded a large panoply of activities' sectors (Hemalakshmi and Muthukumaravel, 2022; Roy and Tamang, 2021; Lamba et al., 2021; Nour et al., 2024). In particular, machine learning (ML) (Emary et al., 2014; Atteia et al., 2023; Azar et al., 2012; Ashfaq et al., 2022; Banu et al., 2017; Aziz et al., 2013; Inbarani and Azar, 2020; Azar et al., 2013) and deep learning (DL) (Koubâa et al., 2020; Jothi et al., 2013; Anter et al., 2013) approaches and software tools are intensively used in various economic fields to achieve both industrial activities, commercial transactions as well as firms management and monitoring.

In this trend, the BP field and the related DSS environments have not been spared by the ubiquitous AI phenomenon. In fact, modern companies operate in open and competitive systems, where surviving in such contexts becomes very tough. To face such challenges, companies have firstly to master their BPs and then to make the suitable decisions at the right time by relying on efficient software tools.

In the following, the BP domain with its underlying concepts and definitions are presented, then DSS related works are exposed. After that, the motivations of this work are explained and the research gap is highlighted.

### 2.1 Presentation of the BPs domain

*Definition 2.1 (BP):* A BP consists of a set of business activities undertaken by one or more organisations in pursuit of some goal (Deutch and Milo, 2007).

As illustrations from the real-world, BPs can range from booking a trip and placing an order to process a retirement application or even managing the procedure of foreign citizens visa. The aforementioned examples of BPs are constructed of a sequence of separate steps, and each one specifies particular business rules that reflects the organisation's business logic. The global dynamic of the BP is governed by a set of constraints (*execution order, time, resources, and so on*) (Khebizi et al., 2017). To handle such constraints and to attempt a smoother and straightforward management and maintenance of those BP's, business process management (BPM) is recognised in the literature as a comprehensive discipline that handles the entire life-cycle of a BP which is articulated on the following four consecutive stages (Tenschert et al., 2020).

- 1 *Modelling:* formal or graphical tools that are both intelligible and executable are frequently used to represent BP models.

- 2 *Implementation*: the activities of the BP are integrated in the information system once they have been developed, documented, and validated.
- 3 *Execution*: the implemented BPs are placed in an execution environment and they must be managed and monitored (Dumas et al., 2023).
- 4 *Monitoring*: finally, after the BP deployment, the gathered data are exploited to examine how the process operates.

To coordinate such important stages using BPM paradigm, a business process management systems (BPMS) (Pranata et al., 2023) should be used as a software program including a collection of appropriate technologies and tools useful to define, automate, and assess the BP's. Indeed, the BP models created in the initial phase of the BPM life-cycle describe the abstract specifications reflecting the business logic supported by the firm's current procedure. The next stages of the BP life-cycle focus on complementary critical stages beneficial for deploying, analysing, monitoring, and optimising the models and their related data in order to ensure the agility of the actual BP.

Many formal and graphical models, ranging from labelled transitions systems (LTS), graphs, Petri-nets, finite state machines, and graphic representations (*UML diagrams*) are used in the literature to represent the BPs specifications. Each of which is endowed with its own specific expressiveness level and captures a collection of constraints such as order, time, and transitional ones (Barroca et al., 2018).

Although this richness of models, selecting the most expressive one is a difficult task for the process manager who operates in a particular context for achieving a specific management goal. Hence, making any decisions regarding the considered BP becomes a complicated and tedious activity. In such scenarios, using automated solutions such as modern chatbots, appears as a promising prospect to support and assist managers in their usual decision-making tasks.

In what follows, the most emergent works having deal with BP decision-making are exposed.

## 2.2 Literature review

To improve the performance of DSS, the junction of NLP and decision-making processes has received a substantial attention in recent research literature by integrating more innovative approaches. The advent of advanced LMs, particularly those based on transformer architecture, has propelled this ambition to new horizons consolidated with more possibilities.

In what follows, the pivotal research works having contributed to the improvement of the monitoring tasks related to decision-making in business processes are discussed. Then, the most relevant works having tackled the issue of incorporating AI techniques in DSS related to BPs will be examined.

### 2.2.1 Smart BPs decision making

Recent approaches trying to incorporate process mining and DL techniques are gaining increasing attention. In fact, the BP management fields have benefits from a panoply of approaches and techniques that involve making decisions based on specific business

rules and resource constraints within a process. Such approaches aim to improve decision-making processes by leveraging large amounts of data and various management rules suitable for advanced analytics. By exploiting and analysing the large amounts of data generated by BPs, such approaches can uncover hidden patterns and insights that may not be apparent through traditional methods.

The data analysis process applies ML algorithms and decision-based approaches. Thus, it can identify patterns (Jan et al., 2019) and measure the variation between the discovered model and the original data (Lee and Verbeek, 2018; van Dongen et al., 2021). Other works focus on making predictions about future outcomes (Di Mauro et al., 2019; Taymouri et al., 2020; Moon et al., 2021; Bukhsh et al., 2021; Gamoura, 2021; Folino et al., 2022) and detecting anomalous behavior (Nolle et al., 2018b,a; Sarno et al., 2020; Krajsic et al., 2021; Elaziz et al., 2023). Finally, the studies conducted by Van der Aalst et al. (2016) have long been instrumental. This work on process mining, consists in extracting insights from event logs, and discovering process models that align with the goal of enhancing decision-making through a data-driven approach.

The deployment of the previous approaches can lead to more accurate decision-making and better overall performance of the BPs management. Overall, decisive-based approaches on BPs have the potential to bring significant benefits to organisations, including better decision-making, increasing efficiency, and improving competitiveness.

### 2.2.2 *Fine-tuned generative models for decision making*

Brown et al. (2020) presented a seminal work on language models as few-shot learners, revealing the extraordinary capacity of models like GPT-3.5-turbo to generalise knowledge from limited examples. This finding laid the foundation for subsequent studies exploring the applicability of such models in complex decision scenarios. In parallel, fine-tuning, as discussed by OpenAI (Aydın and Karaarslan, 2022), emerged as a crucial mechanism for tailoring these large-scale language models to specific domains, providing an avenue for leveraging their inherent language understanding capabilities in targeted contexts, such as decision-making. Furthermore, the work of Shreyashree et al. (2022) introduces the influential language model *bidirectional encoder representations from transformers* (BERT) which emphasises bidirectional contextual embedding. While the focus of this paper lies on GPT-3.5-turbo, the comparative exploration of numerous pre-trained models contributes to a nuanced understanding of the strengths and limitations within the landscape of decision-making models. Understanding the architectural underpinnings of these models is essential for comprehending their capabilities and limitations, especially in the context of decision-making complexity.

The next subsection exposes the research gap consisting of the deployment of chatbots fine-tuned with GPT-3.5-turbo in the DSS context and their integration in the decision information system. Further, it discusses the motivations having led to this work.

### 2.3 *Research gap and motivations*

In what follows, the research gap is highlighted and the motivations are discussed.

### 2.3.1 Research gap

The transformer models presented in Oketunji (n.d.) and Zhang et al. (2024) are considered as the pioneering architecture and the cornerstone upon which GPT-3.5-turbo is built. By integrating fine-tuned GPT-3.5-turbo models into BPs decision-making scenarios, Beheshti and Yang (2023) and Busch et al. (2023) aim to bridge the gap between traditional decision-making approaches and cutting-edge AI technologies. Building upon foundational works in decision theory and AI, the authors suggest a novel approach which leverage chatbots empowered by fine-tuned GPT-3.5-turbo models. These advanced language models are able to understand and generate human-like text, allowing to provide sophisticated insights, recommendations, and support within BP decision contexts.

By harnessing the power of GPT-3.5-turbo, the previous work aims to enhance decision-making tasks in the BPs field, by facilitating natural language interactions, automating routine procedures, and offering intelligent guidance based on the analysis of vast amounts of knowledge. Through this innovative integration, this study target to empower decision-makers with advanced AI capabilities, ultimately improving the efficiency, effectiveness, and adaptability of BP decision-making in complex and dynamic environments. As it was delved into the synthesis of chatbot and GPT-3.5-turbo within the realm of BPs decision-making, the previous works build the groundwork for the current exploration.

Basing upon existing achievements, this paper aims to contribute to the evolving landscape of DSS and bringing forth novel insights based on an approach that leverages advanced LM in the area of BPs.

### 2.3.2 Motivations

The contemporary business landscape is marked by unprecedented complexities and dynamic challenges, requiring organisations to permanently refine their decision-making processes (Aldoseri et al., 2024). In this context, exploiting AI techniques contributes significantly to face the arising challenges. In fact, the fusion of AI and NLP holds immense promise for revolutionising how decisions are made in diverse business domains (Babu and Akshara, 2024). More precisely, this work focuses on exploring and leveraging the capabilities of chatbots and customises the GPT-3.5-turbo model (Marko, 2023) in order to build a new chatbot named '*BP-DSS3*' that empowers and supports decision-making within intricate BPs. The current work is motivated by the following reasons.

- 1 *Handling dynamic evolution of BPs*: Modern companies evolve in open environments where the conducted BPs span across geographical boundaries. Hence, markets are becoming global and more competitive (Clapp, 2023). To cope with the induced effects, firms have to be flexible and able to adapt to the ever changing conditions that occur in their environments at an increasing rate. As an immediate consequence, the intricacies of decision-making have grown exponentially (Saleheen et al., 2024), involving multifaceted considerations spanning strategic planning, resource allocation, risk management, and operational efficiency. To deal with this complex landscape and to manage the associated constraints, there is a compelling need for intelligent systems that can

comprehend, contextualise, and augment the decision-making prowess of human stakeholders. In this perspective, this current work attempts to consolidate DSS with AI techniques.

- 2 *Deploying NLP in the DSS context:* As a sub-field of AI, NLP plays a vital role in DSS by enabling extraction, analysis, and interpretation of information from unstructured textual data. In the DSS context, NLP techniques are employed to process various types of text data, including documents, emails, customer's feedback, and social media posts to provide valuable insights and support decision-making processes.

As an illustration, one key application of NLP in DSS is sentiment analysis, where text data is analysed to determine the sentiment or opinion expressed within it. This analysis can help businesses appreciate customer satisfaction, identify emerging trends, and assess public opinion about products, services, or brands. Hence, sentiment analysis results can inform strategic decisions related to marketing campaigns, product development, and customer service initiatives.

Another important use case of NLP in DSS is information extraction, where relevant information is extracted from text documents or sources to support decision-making processes. This may involve identifying key entities (*such as people, organisations, or locations*), extracting important events or facts, or summarising vast amount of textual data into concise and actionable insights. Thus, information extraction techniques enable DSS to automatically process and analyse vast amounts of textual information, facilitating faster and more informed decision-making.

Furthermore, NLP techniques are utilised in DSS (Al-Amin et al., 2024) for text classification and categorisation tasks related to BPs, where text data is classified into predefined categories or topics. This can help organise and structure textual information, making it easier to navigate and analyse. Text classification algorithms can be applied to numerous domains, such as document management, content filtering, and risk assessment, to streamline decision-making processes and improve efficiency.

Overall, NLP enhances DSS by unlocking valuable insights from unstructured textual data, enabling organisations to make more informed decisions, anticipate trends, and respond effectively to changing business environments. By leveraging NLP techniques, DSS can harness the power of text data to drive strategic initiatives, optimise operations, and gain a competitive edge in today's data-driven landscape.

- 3 *Exploiting chatbots advantages for DSS:* Chatbots with their conversational AI capabilities, offer a valuable opportunity to enhance decision support systems (Al-Amin et al., 2024). By engaging users in natural language conversations, the model can comprehend the subtleties of decision points, consider diverse factors, and provide context-aware recommendations. The potential benefits range from improved decision accuracy to enhanced collaboration among decision-makers, marking a significant advancement of the landscape of intelligent decision support. The major benefits of using chatbots are the following:

- *24/7 availability*: Chatbots enable businesses to provide round-the-clock support to users and customers. Further, they enhance accessibility and responsiveness, especially in global operations.
- *Cost saving*: Tasks' automation leads to cost savings by reducing the need for human intervention. Thus, chatbots can handle a large volume of inquiries simultaneously.
- *Scalability*: Chatbots are easily scalable to handle an increasing number of interactions with numerous users. Furthermore, they are adaptable to the growing needs of the business.
- *Data analysis and insights*: Chatbots can collect and analyse data from interactions, providing valuable insights. They enable businesses to understand user behavior and preferences.
- *Consistency*: Chatbots can ensure consistent responses and adherence to business rules and policies. Consequently, they reduce the likelihood of human errors in communication.

Based on the aforementioned considerations, this work aims to overcome the observed limitations and constraints of traditional DSS and offer decision-makers a sophisticated and perceptive ally by building an enhanced chatbot refined with GPT-3.5-turbo, which serves as a powerful tool during BPs management.

To this end, the next section suggests a framework describing a chatbot that will be used for supporting decision-making related to BPs.

### 3 Development of the chatbot BP-DSS3

Generative language models, such as employed in ChatGPT are designated to generate human-like text that express responses or outputs to users queries, basing on the data introduced in input. In the context of BPs, elaborating a decision-making framework consists of customising the language model for particular use cases, by fine-tuning GPT-3.5-turbo.

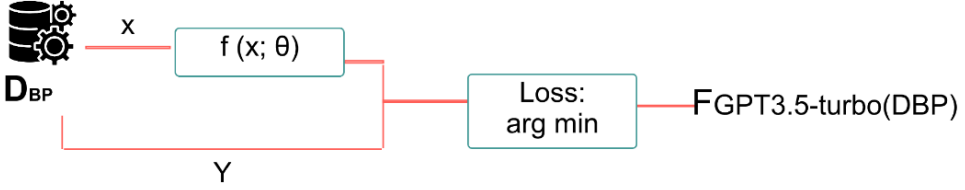
To this end, this section defines the problem being addressed and presents its formal specification, then the different components and stages reflecting the architecture and functionalities of the conceived framework are deeply explained.

#### 3.1 Problem statement and specification

The goal of this work consists to adapt the existing GPT-3.5-turbo model to perform decision-making tasks, such as BP documentation, anomalies detection, or recommendation systems. This goal can be achieved by exploiting BP-specific datasets, through a pragmatic analysis of their content and links. The target model starts learning the domain's specific concepts, patterns, and decision-making logic, as well as the applied business rules. After that, the fine-tuning of GPT-3.5-turbo on BP data involves adjusting the model's pre-trained parameters using a domain-specific dataset. The previous steps allow improving the model's ability to understand and generate text related to BPs, while minimising the loss on data related to BP-specific tasks.

Figure 1 depicts the problem and its parameters.

**Figure 1** Illustration of the BP decision-making problem using chatbots technology (see online version for colours)



Where the used notations are explained below.

- Let  $\mathcal{F}_{\text{GPT3.5-turbo}}(\mathcal{D}_{\text{BP}})$  be the fine-tuned GPT-3.5-turbo model on BP tasks, where  $\mathcal{D}_{\text{BP}}$  designates the domain-specific BP dataset used for the fine-tuning stage.
- $\theta$  corresponds to the model's parameters used in GPT-3.5-turbo.
- Consider that  $f(x; \theta)$  expresses the function producing the output of GPT-3.5-turbo, given the input  $x$  (*contextual BP data*) and the parameters  $\theta$ .
- Now assume that  $\mathcal{L}$  is the loss function capturing the error between the model's output and the true label  $y$  (*ground truth for BP tasks*).
- This previous loss is captured using  $\mathbb{E}_{(x,y) \in \mathcal{D}_{\text{BP}}}$ , which denotes the expectation over the dataset  $\mathcal{D}_{\text{BP}}$  containing the input-output pairs  $(x, y)$  specific to the BP domain.

Basing on the previous specifications, the formal definition of the decision-making problem using chatbots is formalised with the equation (1).

$$\mathcal{F}_{\text{GPT3.5-turbo}}(\mathcal{D}_{\text{BP}}) = \arg \min_{\theta} \mathbb{E}_{(x,y) \in \mathcal{D}_{\text{BP}}} \mathcal{L}(f(x; \theta), y) \quad (1)$$

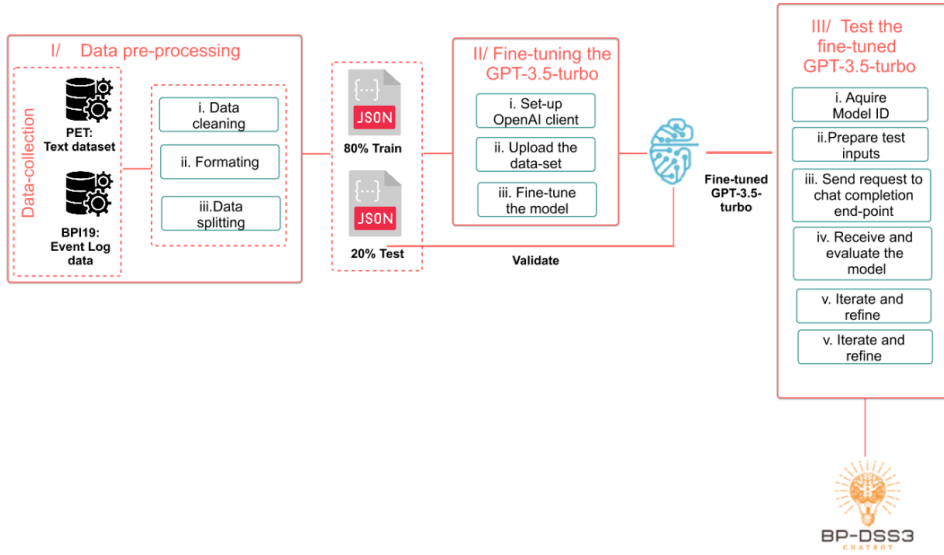
where  $\arg \min$  denotes the habitual mathematical function for calculating the loss between the input and the output values.

This iterative fine-tuning process minimises the loss function  $\mathcal{L}$  and ensures that the target model evolves to a level that it becomes highly specialised for tasks related to the BPs domain.

### 3.2 Development of the chatbot BP-DSS3

This sub-section describes the different modules of the developed chatbot BP-DSS3 and the progress of the processing flow when handling user requests. As shown in Figure 2, BP-DSS3 operates in three main progressive stages leading to decision-making related to BPs. In what follows, these stages are deeply explained and illustrated.

**Figure 2** Architecture and functionalities of the developed chatbot BP-DSS3  
(see online version for colours)



### 3.2.1 Pre-processing stage

Data pre-processing is a critical step to ensure that the employed data-set is consistent, appropriately formatted and ready for the fine-tuning process. The pre-processing stage is articulated around the four following steps detailed.

- 1 **Data collection:** The first step toward an accurate fine-tuning process involves acquiring qualitative and quantitative input datasets. To this end, the annotated dataset, named PET (Bellan et al., 2022) is employed. This textual corpora targets the extraction of BPs specifications from natural languages' texts and unstructured data, expressing processes' descriptions. Further, the widespread event-log data BPI19, capturing various activities performed by a real-life loan application process was intensively exploited. This second dataset is provided by the 9th International Business Process Intelligence Challenge (*ICPM on 2019*) (van Dongen, 2019). Its content consist of a huge event log data reflecting a real-life data over 1,5 million events originating from a Netherlands large multinational company operating in the area of coatings and paints.

Figure 3 depicts an excerpt of BPI19 dataset employed to fine-tune the GPT3 model.

- 2 **Data cleaning:** Data cleaning is a fundamental operation for preparing the dataset to be exploited during the fine-tuning process of GPT-3.5-turbo model. Indeed, cleaning the data helps ensuring that the model receives and manages high-quality input and, consequently, it can generate meaningful and accurate outputs. The two activities involved during data cleaning are the following.

- For the PET text data, data cleaning consists in removing special characters, punctuation, and unnecessary white-spaces of the tokens.
- Identifying and handling missing values for the BPI19 event log dataset. This might involve imputing missing values, removing the columns or records with incomplete data.

**Figure 3** A formatted excerpt of the BPI19 dataset (see online version for colours)

```
{
  "messages": [
    {
      "role": "system",
      "content": "A bot aim to answer the questions about BPI Challenge 2019."
    },
    {
      "role": "user",
      "content": "What BPI challenge 2019?"
    },
    {
      "role": "assistant",
      "content": "This data originated from a large multinational company operating from The Netherlands in the area of con"
    }
  ]
}
```

- 3 *Formatting*: To start the fine-tuning process, the cleaned dataset was prepared for training the developed model. In this perspective, a set of conversations scenarios similar to those that could be asked during real-time conversation sessions was elaborated. Each example in the formatted dataset should be a conversation in the same format as OpenAI's chat completions API. Specifically, conversations' messages are structured as objects with the following attributes.

- *Role*: Can be either 'system', 'user', or 'assistant'. First, a system message is included at the beginning to set the behavior and context for the assistant. This message helps guide the model's responses. Second, user messages are structured in a conversation format, including both user and assistant turns. The use of the role labels aims to differentiate between user and assistant messages.

In this context, the role can take three types of messages in a conversation: system, user, and assistant.

- a The system message comes first and sets the tone for how the assistant should behave.
- b The user messages correspond to the user questions and queries.
- c The assistant messages express the responses provided by the assistant agent.

These labels help keep track of who's saying what in the conversation.

- *Content*: A string containing the current message originating either from the user or the assistant.

- 4 *Data splitting*: The resulting output of the formatting step are stored in a JSON format (Smith, 2015), due to its simplicity and portability. The produced files are splitted into two distinguished parts. The first one contains 80% of data and will be used for the training process, while the remaining 20% of data will serve during tests of the system.

Once the pre-processing step is achieved, data becomes ready to be used for the fine-tuning stage. In what follows, the fine-tuning process is deeply tackled.

### 3.2.2 Fine-tuning stage

To develop the chatbot BP-DSS3, the GPT-3.5-turbo model was used and we were inspired from the work (Tey, 2023). Thus, by using the OpenAI API with the TypeScript (TS) to fine-tune the developed chatbot, the focus was made on the iterative analysis of the system's responses corresponding the users' queries. Hence, the model was trained to understand nuanced decision contexts across different business domains.

**Figure 4** The setup of OpenAI API client code of the GPT3.5-turbo (see online version for colours)

```
import fs from 'fs'
import OpenAI from 'openai'

import { FineTuningJobEvent } from 'openai/resources/fine-tuning'
import 'dotenv/config'

// Gets the API Key from the environment variable `OPENAI_API_KEY`
const client = new OpenAI({
  apiKey: process.env.OPENAI_API_KEY
})
```

**Figure 5** The code for uploading the dataset (see online version for colours)

```
async function main() {
  console.log(`Uploading file`)

  let file = await client.files.create({
    file: fs.createReadStream('./scripts/data.jsonl'),
    purpose: 'fine-tune'
  })
  console.log(`Uploaded file with ID: `, file)

  console.log('-----')

  console.log(`Waiting for file to be processed`)
  while (true) {
    file = await client.files.retrieve(file.id)
    console.log(`File status: ${file.status}`)

    if (file.status === 'processed') {
      break
    } else {
      await new Promise(resolve => setTimeout(resolve, 1000))
    }
  }
}
```

**Figure 6** The code for fine-tuning process of the GPT3.5-turbo model (see online version for colours)

```

console.log(`Starting fine-tuning`)
let fineTune = await client.fineTuning.jobs.create({
  model: 'gpt-3.5-turbo',
  training_file: file.id
})
console.log(`Fine-tuning ID: ${fineTune.id}`)

console.log('-----')

console.log(`Track fine-tuning progress:`)
console.log("status is ", fineTune.status)
const events: Record<string, FineTuningJobEvent> = {}

while (fineTune.status == 'validating_files' || fineTune.status == 'running' ) {
  fineTune = await client.fineTuning.jobs.retrieve(fineTune.id)
  console.log(`${fineTune.status}`)

  const { data } = await client.fineTuning.jobs.listEvents(fineTune.id, {
    limit: 100
  })
  for (const event of data.reverse()) {
    if (event.id in events) continue
    events[event.id] = event
    const timestamp = new Date(event.created_at * 1000)
    console.log(`- ${timestamp.toLocaleTimeString()}: ${event.message}`)
  }

  await new Promise(resolve => setTimeout(resolve, 5000))
}

```

Below, the pivotal steps of the fine-tuning stage using TS with the OpenAI GPT-3.5-turbo API are depicted.

- 1 *Setup OpenAI client:* To setup the OpenAI client for GPT-3.5-turbo, an OpenAI API key is used to authenticate the user requests to the OpenAI API. As illustrated in Figure 4, this action provides specific endpoints for the fine-tuning process.
- 2 *Uploading the dataset:* As already stated in Subsection 3.2.1, fine-tuning a model typically involves handling a dataset that is in the format ('role', 'content'). The code shown in Figure 5 allows uploading the dataset file.
- 3 *Fine-tuning process:* The process is triggered by submitting requests to the OpenAI fine-tuning API. It provides the user the dataset and specify the task or role for which he wants to fine-tune the model. Further, the OpenAI API will guide him through the steps needed for fine-tuning. The code in the Figure 6 illustrates this fine-tuning process.

Once the model has been adequately fine-tuned, it becomes exploitable and its capabilities to address decision-making concerns are examined in the following.

### 3.2.3 Testing the fine-tuned model

The testing procedure, typically, involves formulating and submitting the users' requests to the chat completion endpoint with the specific ID of the fine-tuned model. The rightmost part of the Figure 2 enumerate the different activities to be conducted during the test procedure. Hereafter the considered activities are deeply explained.

- 1 *Acquire model ID*: Consists in obtaining the unique identifier (ID) assigned to the fine-tuned GPT-3.5-turbo model (e.g., *in this case the ID is* ft:gpt-3.5-turbo-0613:university-of-8may::8fChoSKr). This ID is necessary for accessing and utilising the model for inference tasks related to decision-making in the BP context.
- 2 *Prepare test inputs*: An adequate set of test inputs representing different conversational scenarios is prepared and ready to be introduced as input of the developed chatbot BP-DSS3. These inputs concern decision-making on BPs and should cover a diverse range of contexts, use-case scenarios and potential user queries or prompts.
- 3 *Send requests to chat completion endpoint*: This task utilises the OpenAI API (*or any other appropriate interface*) to send requests to the chat completion endpoint of BP-DSS3. Each request must include the ID of the concerned fine-tuned model along with the related test inputs.

The three previous steps (1, 2 and 3) are shown in the third box (No. 3) of Figure 2 and their implementation is illustrated in Figure 7. At this development stage, the evaluation and refinement of the resulting responses of the fine-tuned GPT-3.5-turbo model constitute the goal of the three following steps.

- 4 *Receive and evaluate the model responses*: Upon receiving the requests, the chat completion endpoint processes the introduced inputs. As a response to the submitted queries, the model generates the corresponding answers based on the provided context and its decision-making capabilities. After that, the generated responses are evaluated by the fine-tuned model accordingly to a set of predefined criteria and metrics. This evaluation consists in assessing the accuracy of the made decisions, the relevance of the responses to the considered context, and the overall coherence and quality of the dialogue between the user and the chatbot BP-DSS3.
- 5 *Iterate and refine*: Based on the results obtained during the evaluation phase, potential situations needing more improvements can be identified and analysed. This may include refining the fine-tuning process, incorporating feedback, adjusting model parameters, or incorporating additional training data to address any shortcomings observed during the previous evaluation phase.
- 6 *Repeat testing as needed*: Here, the testing process is iterated and the necessary adjustments are made in order to enhance the model abilities, while retesting it to validate the operated improvements. This iterative cycle is repeated until the model achieves satisfactory performance according to the desired criteria.

Although the previous steps can effectively test the output sequences provided by the fine-tuned GPT-3.5-turbo model for decision-making, it is important to assess that the developed chatbot BP-DSS3 also meets the standards of performances. In this

perspective, the experiments of the fine-tuned model will allow assessing this important aspect. The next section is dedicated to the discussion of the experimental results.

**Figure 7** The code for utilising the fine-tuned GPT-3.5-turbo model (see online version for colours)

```
const config = new Configuration({
  apiKey: process.env.OPENAI_API_KEY,
});

const openai = new OpenAIApi(config);

export const runtime = "edge";

export async function POST(req: Request) {
  const { messages } = await req.json();

  const response = await openai.createChatCompletion({
    model: 'ft:gpt-3.5-turbo-0613:university-of-8may::8fChoSKr',
    max_tokens: 20,
    stream: true,
    messages: messages,
  });

  const stream = OpenAIStream(response, {});
  openai.createFineTune

  return new StreamingTextResponse(stream);
}
```

## 4 Experimental results

This section aims to experiment the applicability of the proposed chatbot and to assess its performance for supporting decision-making in the context of BPs. The first goal consists to measure and validate the quality of the training stage of the proposed BP-DSS3 model. To this end, the efficiency of the trained model when handling the datasets used during the fine-tuning process is examined. After that, the abilities of the developed chatbot is experimented with two real-world scenarios [i.e., *alignment with organisational objectives (AOOs)* and *risk management and contingency planning (RMCP)*] and the obtained results are illustrated and discussed. Finally, the potential challenges and issues related to the large adoption of chatbots in DSS environments are discussed in order to mark out new research perspectives.

Before going into further experimental details, we recall that the employed datasets during the training stage are:

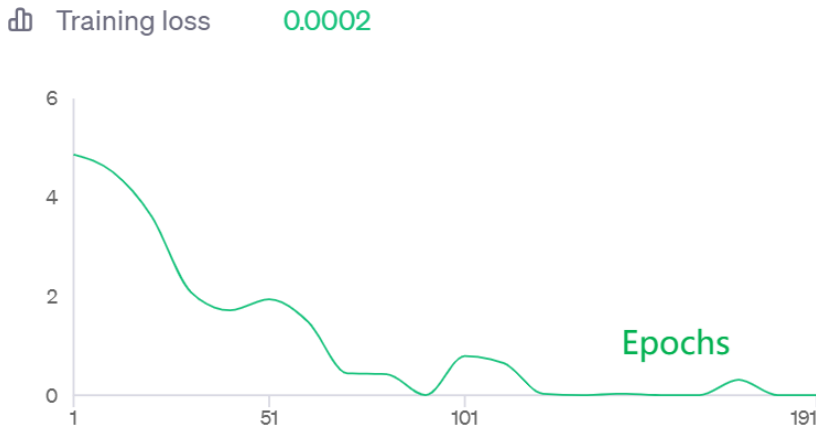
- 1 The PET dataset which is used to extract BP models from natural languages.
- 2 The BPI19 event-log data containing real-word BPs execution. For more details about the used fine-tuning datasets, see the previous first item (data collection) in Subsection 3.2.1.

The following subsection aims to validate the decision-making capabilities of the proposed system through the training stage.

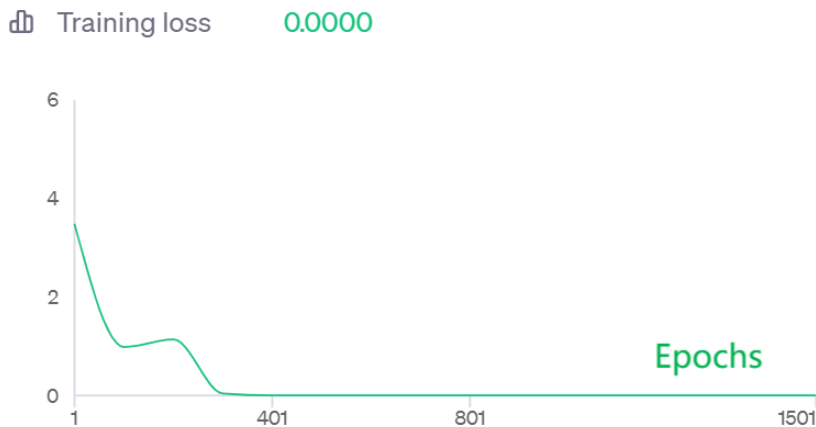
#### 4.1 Capabilities of BP-DSS3 for decision-making

These experiments aim to showcase the abilities of the developed chatbot BP-DSS3 for supporting decision-making tasks during its training on the pre-processed two datasets PET and BPI19. To this end, the chatbot is trained with the two previous datasets (i.e., *PET* and *BPI19*) and for each one the training loss is estimated.

**Figure 8** The GPT-3.5-turbo trained with PET dataset (see online version for colours)



**Figure 9** The GPT-3.5-turbo trained with BPI19 dataset (see online version for colours)



The training results of the model for more than 100 epochs with the two datasets are shown in Figures 8 and 9, respectively.

The interpretation of the two figures brings out the two following conclusions.

- 1 The accuracy of the training phase of both datasets is higher than 99% (*0.9998 for the PET dataset and 100% for BPI19 dataset*). During the fine-tuning of GPT-3.5-turbo on a specific BP dataset, the model undergoes a rapid training process due to the sophisticated pre-trained architecture that already encompasses a vast amount of generalised knowledge. The model does not require the same degree of training complexity as traditional neural networks which act from scratch. As a result, the fine-tuning stage is relatively fast, depending on the dataset size and the specific domain-related adjustments. This efficiency allows the model to adapt quickly to new contexts, while preserving the high accuracy of its predictions. Even during rapid training phases, GPT-3.5-turbo can achieve near-perfect adaptation to the specific domain without prolonged training epochs.
- 2 Further, such training results prove the efficiency of the proposed framework and its capabilities to learn patterns and associations between input messages and appropriate responses. This fine-tuning of GPT-3.5-turbo on BP datasets like BPI19 and PET uncovers a rich array of patterns, from sequential workflows to conditional decision-making and exception handling. These patterns help the model to become highly contextualised and customised for BP-related tasks, leading to more accurate predictions, better anomaly detection, and a deeper understanding of process variations. By understanding the previous patterns, organisations can use the model to automate decision-making (e.g., *identifying inefficiencies or areas*) for improvement in their BPs.

The following subsection describes a more practical use of the developed chatbot BP-DSS3 and tests its effectiveness for managing handling two real-world scenarios.

#### 4.2 Case studies

The goal of deploying the developed chatbot for further case studies aims to experiment its applicability and its efficiency for handling real-world BPs. Afterwards, the obtained results are compared with those provided by ChatGPT before the fine-tuning stage. More precisely, the experimental objectives of this stage are the following.

- 1 Whatever the accuracy (Varoquaux and Colliot, 2023) of BP-DSS3 is very satisfying, this parameter alone may not fully capture the chatbot's decision-making capabilities. Thus, resorting to the usage of real-world scenarios can consolidate its performances' features during its effective deployment.
- 2 Generally, over-fitting is a common concern during the training phase of ML models. In this context, the over-fitting issue does not pose significant difficulties, even with high training accuracies (*99% or more*). This observation is largely justified by the GPT-3.5's ability to generalise knowledge effectively, due to its massive pre-training on diverse datasets. As a consequence, the experimental results that are based on testing real cases will examine the absence of the over-fitting phenomena.

The developed chatbot BP-DSS3 was experimented in the field of strategic planning. Strategic planning ensures that BP decision-making aligns with the overall objectives and missions of the organisation. This involves identifying key strategic priorities, such as increasing profitability, improving customer satisfaction, or expanding market share, and developing BPs that support these priorities. By aligning decision-making with strategic objectives, organisations can optimise their resources and focus their efforts on activities that drive long-term success.

For experimental requirements, a set of BPs aimed at setting goals have been identified and consisely expressed. Than, long-term strategies, as well as the outlining needed actions to achieve those goals are formulated and implemented. After that, the developed chatbot BP-DSS3 is deployed for assisting users in decision-making tasks. For space reasons, the experimental results are illustrated basing only on two important BPs related to strategic planning. The first BP concerns the AOOs, while the second one focuses on RMCP.

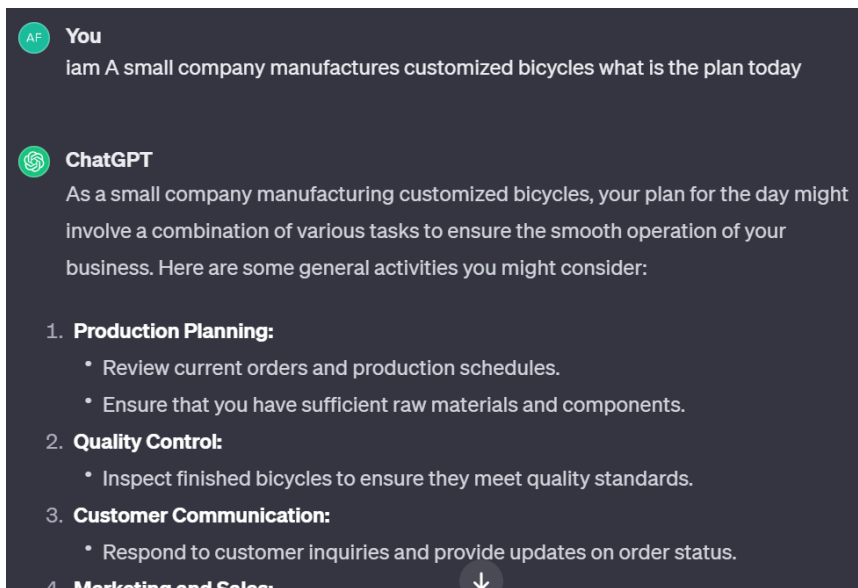
For each BP, relevant requests have been formulated, and submitted to both chatbots, i.e.; chatGPT and BP-DSS3. Finally, the output sequences are analysed and compared.

Hereafter, the conversations scenarios generated during dialogue sessions between users and the two chatbots are depicted and discussed.

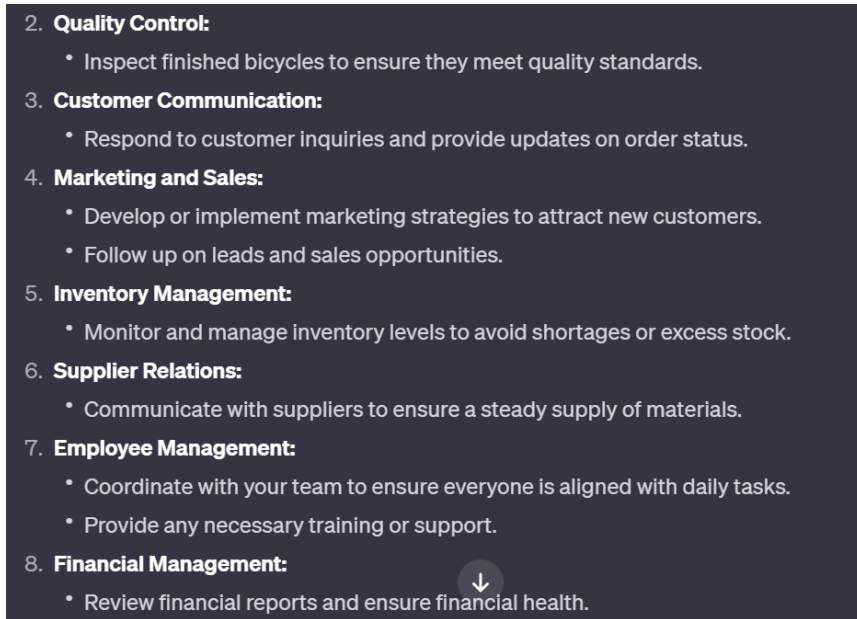
#### 4.2.1 Scenario 1: AOO

As observed in Figures 10, 11 and 12, the answers returned by the general ChatGPT for AOO concerns are general and very vague. Consequently, they cannot satisfy the decision-making needs in an effective way.

**Figure 10** Responses of the general ChatGPT to questions of the AOO BP (part 1)  
(see online version for colours)



**Figure 11** Responses of the general ChatGPT to questions of the AOO BP (part 2)  
(see online version for colours)



**Figure 12** Responses of the general ChatGPT to questions of the AOO BP (part 3)  
(see online version for colours)



However, after the fine-tuning process, the developed chatbot BP-DSS3 provides more accurate answers that can be exploited immediately by firms managers. As illustrated in Figures 13 and 14, the exact business objectives are concisely considered by the developed BP-DSS3 chatbot, which supports the managers making informed decisions during their daily activities. In fact, this chatbot can facilitate strategy planning by simulating different business scenarios, relying on historical knowledge of market trends.

This capability helps business leaders anticipate potential challenges related to threats and opportunities, and consequently develop robust strategies.

**Figure 13** Response of the BP-DSS3 chatbot for the AOO BP (part 1) (see online version for colours)



Hello, I am A small company manufactures customized bicycles



Hello, I am A small company manufactures customized bicycles.



what should i do as a Plans after the department of the bicycles receive an order



After receiving an order, an authorized clerk checks the order and creates an order confirmation.



Then what is the Second step after a new process instance is created?



After the creation of the order confirmation, the sales department starts negotiating a delivery date with the customer.

**Figure 14** Response of the BP-DSS3 chatbot for the AOO BP (part 2) (see online version for colours)



if A member of the sales department reject



If the sales department rejects the order, the process is finished.



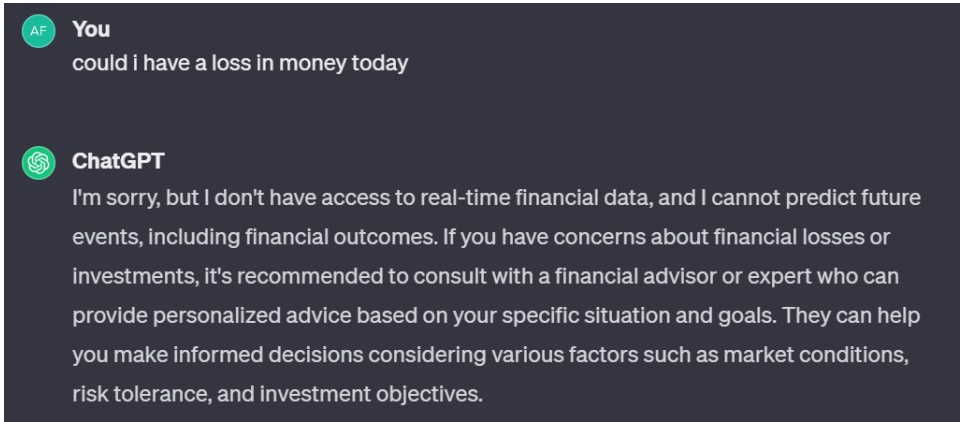
if A member of the sales department accept



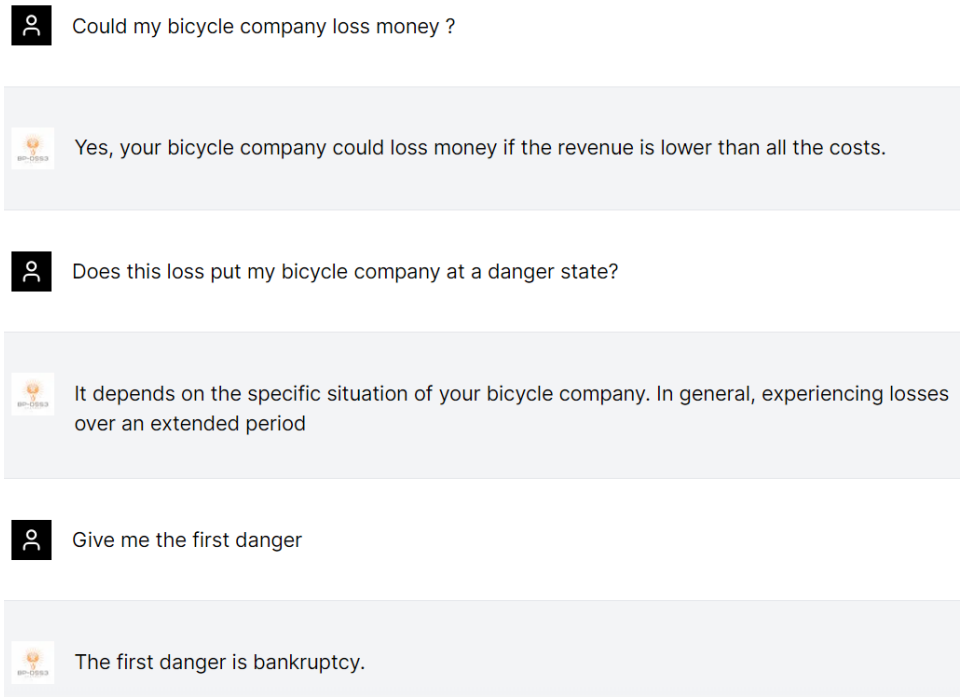
If the sales department accepts the order, the clerk starts the order fulfilment.

The following case study is dedicated to the evaluation of the second BP that expresses strategic planning related to RMCP.

**Figure 15** Response of the general ChatGPT to suggest the possible future risks (see online version for colours)



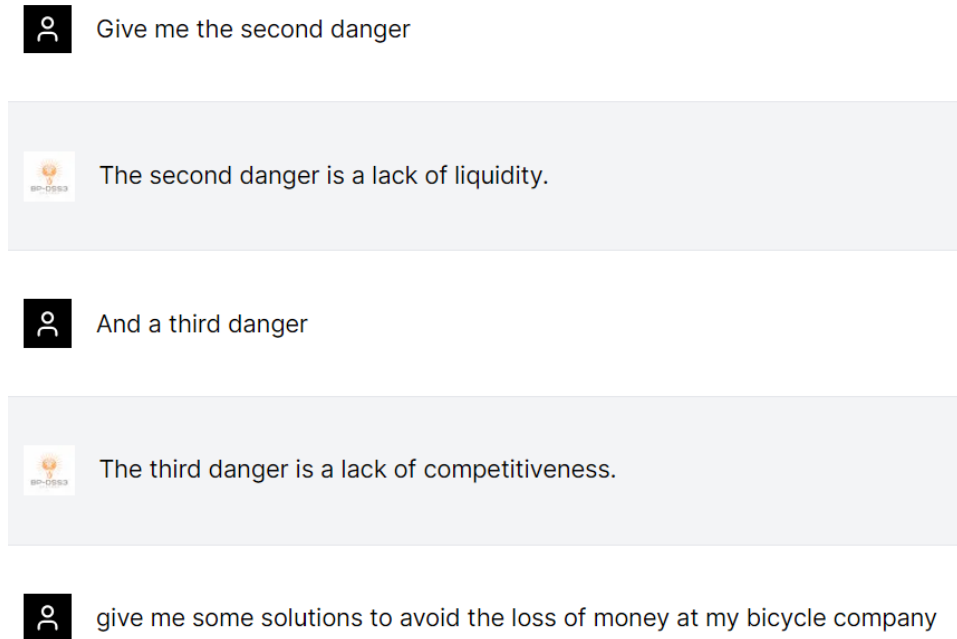
**Figure 16** Response of the BP-DSS3 chatbot for predicting future risks (part 1) (see online version for colours)



#### 4.2.2 Scenario 2: RMCP

Effective strategic planning involves identifying and attenuating risks associated with BP decision-making. This includes assessing potential risks and uncertainties, such as market volatility, regulatory changes, or technology disruptions, and developing contingency plans to address such challenges. By proactively managing risks, organisations can minimise the impact of unforeseen events on their operations and maintain business continuity.

**Figure 17** Response of the BP-DSS3 chatbot for predicting future risks (part 2)  
(see online version for colours)



The response of the general ChatGPT to the particular concern related to *RMCP*: “could I have a loss of money today?” is shown in Figure 15. It is observed that ChatGPT is unable to handle such a question and it advises the user to consult a financial expert.

However, the same previous issue was effectively handled and resolved with the developed BP-DSS3, which can identify the potential risks and their impact on the company success and efficiency. As illustrated in Figures 16 and 17, the chatbot BP-DSS3 exhibits remarkable skills in handling firms potential risks and for addressing contingency plans.

In summary, the proposed BP-DSS3 provides more detailed and tailored replies to the target domain concerns, comparing with the general ChatGPT.

To assess a model performances, such as BP-DSS3, a set of known metrics is used in order to compare the model’s predictions to actual labels. In this regard, the previous test case studies are evaluated in the following subsection through four performance metrics, namely: *precision* (Manning, 2008), *recall* (Manning, 2008), *f1.score* (Manning, 2008) and *accuracy* (Varoquaux and Colliot, 2023).

### 4.2.3 Performance metrics evaluation

In the general context of a chat-bot based decision-making system, such as BP-DSS3, the performance score must indicate the abilities of the model to effectively perform the users' requests. Although the developed system can be deployed and exploited for decision-making purposes, it would be appropriate to analyse its real performances in order to further improve its capabilities. In this perspective, the confusion matrix (Townsend, 1971) serves as a diagnostic useful tool which highlights the distribution of correct and confused predictions generated by the model. The set of evaluation values contained in this matrix allows stakeholders appreciate the capabilities and weakness of the proposed model. Furthermore, the conventional metrics, such as precision, recall, and F1-score can provide deeper insights regarding the model's reasoning abilities. In this regard, the chatbot's responses resulting from the prior two case studies (i.e., *AOO* and *RMCP*) were exploited to evaluate the BP-DSS3's performances.

As shown in Table 1, the accuracy matrix values for the BP-DSS3 model are measured and compared with those of ChatGPT in order to assess the predicted values against actual values for the two used datasets.

**Table 1** The confusion matrix of BP-DSS3 and ChatGPT

| <i>BP-DSS3</i> |       | <i>Predicted</i> |        | <i>ChatGPT</i> |       | <i>Predicted</i> |        |
|----------------|-------|------------------|--------|----------------|-------|------------------|--------|
| Actual         |       | True             | False  | Actual         |       | True             | False  |
|                | True  | 8 (TP)           | 2 (FN) |                | True  | 1 (TP)           | 9 (FN) |
|                | False | 0 (FP)           | 2 (TN) |                | False | 2 (FP)           | 0 (TN) |

It is important to clarify that in the particular context of BP decision-making, the explicit signification of the values of the different metrics contained in Table 1 are the following.

#### 1 Actual:

- *True*: Expresses the questions derived from the test cases contained in (*BPII9* or *PET*).
- *False*: corresponds to the questions provided outside of the dataset's test cases (*PET* or *BPII9*)

#### 2 Performance metrics:

- *TP*: Represents the number of correctly detected responses from the test cases contained in (*BPII9* or *PET*).
- *TN*: Represents the number of correctly detected responses of the questions provided outside of the dataset's test cases (*PET* or *BPII9*).
- *FP (type I error)*: Expresses the number of responses detected by the chatbot as correct for the questions provided outside of the dataset's test cases (*PET* or *BPII9*).
- *FN (type II error)*: Expresses the number of undetected responses by the chatbot for the questions provided inside of the dataset's test cases (*PET* or *BPII9*).

The analysis of the results displayed in the previous table allows establishing the following interpretations of the performances achieved by the chatbots BP-DSS3's and ChatGPT.

- *True positive (TP)*: BP-DSS3 correctly predicted eight positive instances (i.e., it successfully identified eight correct results), while ChatGPT correctly predicted only one positive instance (i.e., it recognised just one correct positive result).
- *True negative (TN)*: BP-DSS3 correctly identified two negative instances, while ChatGPT did not correctly identify any negative instances.
- *False positive (FP)*: BP-DSS3 did not make any FP errors (i.e., it did not mistakenly detect any negative instances as positive), while ChatGPT detects two cases.
- *False negative (FN)*: BP-DSS3 failed to predict two positive instances (i.e., it misclassified two positive cases as negative), while ChatGPT misclassified nine positive instances as negative, showing a significant number of missed positive predictions.

Basing on the previous interpretations, it appears clearly that the developed chatbot BP-DSS3 outperforms ChatGPT, particularly, in correctly identifying positive cases. In fact, ChatGPT missed a large number of positive instances (nine FNs) and also incorrectly flagged two negative instances as positive, demonstrating areas where its performances could be improved.

**Table 2** A performance evaluation of the test case studies

| Metrics | Precision | Recall | F1.score | Accuracy |
|---------|-----------|--------|----------|----------|
| ChatGPT | 0.33      | 0.1    | 0.09     | 0.08     |
| BP-DSS3 | 1         | 0.8    | 0.89     | 0.83     |

The advanced exploitation of the results obtained in table 1 allows using more accurate evaluation metrics. In fact, four performance metrics (*precision*, *recall*, *f1.score* and *accuracy*) are calculated to understand the strengths and the weaknesses of the chat-based system. These action helps assessing the overall effectiveness of each system in handling test case studies.

Table 2 summarises the used four performance measures and their interpretation is the following.

- 1 *Precision*: The BP-DSS3 have a precision of 1, which indicates that when it predicted something as positive, it was always correct (*no FPs*). However, ChatGPT had a low precision, meaning only 33% of the instances it predicted as positive were actually correct.
- 2 *Recall*: ChatGPT's recall is 0.1, meaning it identified only 10% of the actual positive cases. It missed a substantial number of positive cases (high FNs). On the other hand, BP-DSS3's recall is 0.8, meaning it successfully identified 80% of the positive cases, performing much better than ChatGPT in capturing TPs.

- 3 *F1-score*: ChatGPT's F1-score is 0.15, indicating very poor performance due to its low recall. This means that ChatGPT struggles with both correctly identifying and predicting positive cases. But, BP-DSS3's F1-score is 0.89, meaning it performs much better in balancing both precision and recall. It effectively identifies most of the positive cases with few errors.
- 4 *Accuracy*: ChatGPT's accuracy is 0.08, meaning it correctly classified only 8% of all instances. This highlights ChatGPT's low overall performance in predicting the correct outcomes. In the contrary, BP-DSS3's accuracy is 0.83, meaning it correctly classified 83% of all instances, which is much higher and indicates stronger overall performance.

From the above interpretation, the BP-DSS3 is clearly the superior system in this evaluation, achieving high performance in identifying and correctly predicting the results of the test cases, whereas ChatGPT suffers from low recall and overall accuracy. Overall, BP-DSS3 demonstrates high performance in BP decision-making, while ChatGPT requires substantial improvements to be considered effective in this area. In this context, the following conclusions are achieved:

- BP-DSS3 significantly outperforms ChatGPT in all key performance metrics (*precision, recall, F1-score, and accuracy*).
- BP-DSS3's perfect precision (1.0) and high recall (0.8) suggest it is highly reliable in making correct predictions with very few errors, while ChatGPT struggles with both FPs and FNs.
- ChatGPT's low recall (0.1) and F1-score (0.09) indicate that it misses a large number of positive cases and fails to achieve a good balance between precision and recall.

After validating the proposed BP-DSS3 model, in the next subsection we will discuss and highlight the system's limitations and challenges.

### 4.3 Discussion

The previous experiments involving real-BP datasets proves that the BP-DSS3 model's adaptability are very satisfactory. In fact, the training step reinforced with the fine-tuning mechanism allow the conceived chatbot to avoid memorising specific data points, while retaining the ability to generalise well across unseen previous data. Hence, it provides robust results even with fast training cycles. Thus, the experiment confirms the model's resilience against over-fitting, regardless of the high accuracy achieved during the fine-tuning stage.

After having illustrated the performance of the proposed chatbot and its abilities to handle decision-support concerns, in what follows the limitations and challenges inherent to BP-DSS3 chatbots' are discussed.

- 1 *Model explainability*: The explainability of chatbots, and AI generative models in general, pertains to the ability to understand and interpret the decisions made by the model. This allows the chatbot gaining the user confidence and helps him make good and fast decision, and furthermore preserve the transparency. In fact,

while the chatbot can generate responses that are contextually relevant and coherent, understanding how and why it arrives at a particular response can be challenging. This lack of transparency can be problematic, especially in sensitive or high-stakes applications where users need to trust the reasoning behind the model's decisions.

- 2 *Monitoring and maintenance:* Regular monitoring and updating are needed to ensure that the chatbot performs as expected and generates pertinent results. Thus, the utilisation and maintenance of the chatbot for a long time is required to handle refreshed information and features.
- 3 *Handling complex scenarios:* In numerous scenarios, the chatbots are limited in handling highly complex or ambiguous situations. In this context, human intervention may be necessary for operating particular complex decision-making activities.
- 4 *User experience insights and judgements:* User judgment plays a crucial role in fine-tuning GPT-3.5-turbo for chatbot applications in the BP decision-making field. Such user contribution helps refining the model's responses to ensure that they are contextually relevant and aligned with the organisation's objectives. By incorporating user feedback, the fine-tuning process becomes more adaptive, leading to improved accuracy, trustworthiness, and decision quality in BP scenarios over a long time utilisation.

Additionally, the following ethical considerations of the BP-DSS3 chatbot are considered as potential issues and challenges to be addressed by different stakeholders in the field.

- 1 *Potential biases persist:* Algorithmic biases and limitations in language understanding further exacerbate the explainability issue, resulting often in misinterpretations or inappropriate responses. Moreover, feedback loops can reinforce and amplify existing biases over time. Addressing this challenge requires a holistic approach, including diverse and representative training data, bias detection and attenuation techniques, human oversight and intervention, as well as transparency measures to ensure efficient chatbot interactions.
- 2 *Data privacy and security:* Handling sensitive information requires robust and adequate security policies. Thus, compliance with data protection regulations is crucial.
- 3 *Transparency:* Maintaining transparency about the capabilities and limitations of the chatbot. Users should be aware that they are interacting with an AI system and not a human, which helps set appropriate expectations regarding the quality and reliability of the advice provided.
- 4 *Accountability:* Establishing clear lines of accountability for decisions made based on the chatbot's recommendations. Organisations should ensure that there is a human oversight mechanism to review and validate critical decisions to avoid disasters leading to loss of resources and human lives due to the systematic use of automatic responses.
- 5 *User autonomy:* Respecting user autonomy by encouraging individuals to make their own decisions rather than overly relying on the chatbot's recommendations.

Providing options for users to seek human assistance when needed can foster a more balanced decision-making process.

In this circumstance, the chatbot BP-DSS3 can be maintained and updated over time by periodically fine-tuning GPT-3.5-turbo with new BP data, ensuring that the model adapts to evolving decision-making needs and trends. Continuous monitoring and incorporating feedback loops from real-world interactions can further enhance its accuracy and relevance in BP decision-making.

## 5 Conclusions

Effective and efficient decision-making is at the heart of successful business activities. The integration of advanced technologies, such as NLP, offers novel opportunities to enhance decision support systems. This paper fine-tunes the GPT-3.5-turbo language model to develop a specialised chatbot BP-DSS3, which can perform decision-making support in the context of BPs management. After the training step on adequate data, the BP-DSS3 is deployed for real-world scenarios. The first experimental results demonstrate the efficiency of BP-DSS3 compared to the general ChatGPT for decision-making tasks related to BPs.

The following issues outline the future research directions.

- 1 Connecting the BP-DSS3 with existing business databases and systems in order to validate the achieved finding with more case studies.
- 2 Including a self-learning feature process in the chatbot to adjust and adapt the system capabilities over time through iterative training and refinement procedures.
- 3 Further, the ultimate objective is to include the BPs modelling in the BP-DSS3, which will enable users to access the company's general process abstract model whenever needed.
- 4 Moreover, we plan to integrate the user judgment in the conceived system.
- 5 Finally, expanding future work to consider scalability, adaptability to various industries, and integration with other business intelligence tools would strengthen the development potential of chatbots.

## References

- Al-Amin, M., Ali, M.S., Salam, A., Khan, A., Ali, A., Ullah, A., Alam, M.N. and Chowdhury, S.K. (2024) *History of Generative Artificial Intelligence (AI) Chatbots: Past, Present, and Future Development*, arXiv preprint arXiv:2402.05122.
- Aldoseri, A., Al-Khalifa, K. and Hamouda, A.M. (2024) 'AI-powered innovation in digital transformation: key pillars and industry impact', *Sustainability*, Vol. 16, No. 5, p.1790.
- Anter, A.M., Azar, A.T., Hassanien, A.E., El-Bendary, N. and ElSoud, M.A. (2013) 'Automatic computer aided segmentation for liver and hepatic lesions using hybrid segmentations techniques', in *2013 Federated Conference on Computer Science and Information Systems*, IEEE, pp.193–198.
- Ashfaq, T., Khalid, R., Yahaya, A.S., Aslam, S., Azar, A.T., Alsafari, S. and Hameed, I.A. (2022) 'A machine learning and blockchain based efficient fraud detection mechanism', *Sensors*, Vol. 22, No. 19, p.7162.
- Atteia, G., El-Kenawy, E-S.M., Samee, N.A., Jamjoom, M.M., Ibrahim, A., Abdelhamid, A.A., Azar, A.T., Khodadadi, N., Ghanem, R.A. and Shams, M.Y. (2023) 'Adaptive dynamic dipper throated optimization for feature selection in medical data', *Computers, Materials & Continua*, Vol. 75, No. 1, pp.1883–1900.
- Aydin, Ö. and Karaarslan, E. (2022) *OpenAI ChatGPT Generated Literature Review: Digital Twin in Healthcare*, SSRN: 4308687.
- Azar, A.T., Member, I., Hassanien, A.E. and Kim, T-h. (2012) 'Expert system based on neural-fuzzy rules for thyroid diseases diagnosis', in *International Conference on Bio-Science and Bio-Technology*, Springer, pp.94–105.
- Azar, A.T., Banu, P.N. and Inbarani, H.H. (2013) 'Psorr-an unsupervised feature selection technique for fetal heart rate', in *2013 5th International Conference on Modelling, Identification and Control (ICMIC)*, IEEE, pp.60–65.
- Aziz, A.S.A., Azar, A.T., Salama, M.A., Hassanien, A.E. and Hanafy, S.E-O. (2013) 'Genetic algorithm with different feature selection techniques for anomaly detectors generation', in *2013 Federated Conference on Computer Science and Information Systems*, IEEE, pp.769–774.
- Babu, C.S. and Akshara, P. (2024) 'Revolutionizing conversational AI: unleashing the power of ChatGPT-based applications in generative AI and natural language processing', in *Advanced Applications of Generative AI and Natural Language Processing Models*, pp.228–248, IGI Global.
- Banu, P.N., Azar, A.T. and Inbarani, H.H. (2017) 'Fuzzy firefly clustering for tumour and cancer analysis', *International Journal of Modelling, Identification and Control*, Vol. 27, No. 2, pp.92–103.
- Barroca, B., Amaral, V. and Buchs, D. (2018) 'Semantic languages for developing correct language translations', *SQJ*, Vol. 26, pp.417–453.
- Beheshti, A. and Yang, A. (2023) 'ProcessGPT: transforming business process management with generative artificial intelligence', in *2023 IEEE ICWS*, pp.731–739.
- Bellan, P., van der Aa, H., Dragoni, M., Ghidini, C. and Ponzetto, S.P. (2022) 'PET: an annotated dataset for process extraction from natural language text tasks', in *ICBPM*, pp.315–321.
- Bharadiya, J.P. (2023) 'Machine learning and AI in business intelligence: trends and opportunities', *International Journal of Computer*, Vol. 48, No. 1, pp.123–134.
- Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. and Dhariwal, P. (2020) 333 *Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Sskell, et al. Language Models are 334 Few-Shot Learners*, arXiv preprint arXiv:2005.14165.
- Bukhsh et al. (2021) *ProcessTransformer: Predictive BP Monitoring with Transformer Network*, arXiv preprint arXiv:2104.00721.

- Busch, K., Rochlitzer, A., Sola, D. and Leopold, H. (2023) 'Just tell me: prompt engineering in business process management', in *International Conference on Business Process Modeling, Development and Support*, Springer, pp.3–11.
- Clapp, J. (2023) 'Concentration and crises: exploring the deep roots of vulnerability in the global industrial food system', *The Journal of Peasant Studies*, Vol. 50, No. 1, pp.1–25.
- Deutch, D. and Milo, T. (2007) 'Querying structural and behavioral properties of business processes', in *International Symposium on Database Programming Languages*, pp.169–185, Springer, Berlin, Heidelberg.
- Di Mauro, N. et al. (2019) 'Activity prediction of business process instances with inception CNN models', in *XVIIIth International Conference of the Italian AAI*, Springer, pp.348–361.
- Dumas, M.F. et al. (2023) 'AI-augmented business process management systems: a research manifesto', *ACM TMIS*, Vol. 14, No. 1, pp.1–19.
- Elaziz et al. (2023) 'Deep reinforcement learning for data-efficient weakly supervised business process anomaly detection', *Journal of Big Data*, Vol. 10, No. 1, p.33.
- Emary, E., Zawbaa, H.M., Hassanien, A.E., Schaefer, G. and Azar, A.T. (2014) 'Retinal blood vessel segmentation using bee colony optimisation and pattern search', in *2014 International Joint Conference on Neural Networks*, IEEE, pp.1001–1006.
- Folino et al. (2022) 'Semi-supervised discovery of DNN-based outcome predictors from scarcely-labeled process logs', *Business & ISE*, Vol. 64, No. 6, pp.729–749.
- Gamoura, S.C. (2021) 'Predictive reinforcement learning algorithm for unstructured bp optimisation: case of human resources process', *IJSTDS*, Vol. 1, No. 2, pp.184–214.
- Hemalakshmi, K. and Muthukumaravel, A. (2022) 'An efficient filtering technique for detecting traffic surveillance in intelligent transportation systems', *International Journal of Intelligent Engineering Informatics*, Vol. 10, No. 6, pp.504–521.
- Inbarani, H.H. and Azar, A.T. (2020) 'Leukemia image segmentation using a hybrid histogram-based soft covering rough k-means clustering algorithm', *Electronics*, Vol 9, No. 1, p.188.
- Jan, B., Farman, H., Khan, M., Imran, M., Islam, I.U., Ahmad, A., Ali, S. and Jeon, G. (2019) 'Deep learning in big data analytics: a comparative study', *Computers & Electrical Engineering*, May, Vol. 75, pp.275–287.
- Jothi, G., Inbarani, H.H. and Azar, A.T. (2013) 'Hybrid tolerance rough set: PSO based supervised feature selection for digital mammogram images', *International Journal of Fuzzy System Applications*, Vol. 3, No. 4, pp.15–30.
- Jusman, I.A., Ausat, A.M.A. and Sumarna, A. (2023) 'Application of chatgpt in business management and strategic decision making', *Jurnal Minfo Polgan*, Vol. 12, No. 2, pp.1688–1697.
- Khebizi, A., Seridi-Bouchelaghem, H., Benatallah, B. and Toumani, F. (2017) 'A declarative language to support dynamic evolution of web service business protocols', *Serv. Oriented Comput. Appl.*, Vol. 11, No. 2, pp.163–181, <https://doi.org/10.1007/s11761-016-0204-7>.
- Koubâa, A., Ammar, A., Alahdab, M., Kanhouh, A. and Azar, A.T. (2020) 'DeepBrain: experimental evaluation of cloud-based computation offloading and edge computing in the internet-of-drones for deep learning applications', *Sensors*, Vol. 20, No. 18, p.5240.
- Krajsic et al. (2021) 'Variational autoencoder for anomaly detection in event data in online process mining', in *ICEIS*, Vol. 1, pp.567–574.
- Lamba, M., Munjal, G. and Gigras, Y. (2021) 'A MCDM-based performance of classification algorithms in breast cancer prediction for imbalanced datasets', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 5, pp.425–454.
- Lee, W.L.J. and Verbeek, E.A. (2018) 'Recomposing conformance: closing the circle on decomposed alignment-based conformance checking in process mining', *Information Sciences*, Vol. 466, pp.55–91.

- Manning, C.D. (2008) *Introduction to Information Retrieval*, Vol. 39, pp.234–265, Cambridge University Press, Cambridge.
- Marko, K. (2023) *Applying Generative AI and Large Language Models in Business Applications*, Bachelor's thesis, Bachelor's Programme in Industrial Engineering and Management, Lappeenranta-Lahti University of Technology LUT.
- Moon et al. (2021) 'Pop-on: prediction of process using one-way language model based on nlp approach', *Applied Sciences*, Vol. 11, No. 2, p.864.
- Nolle, T., Luetttgen, S., Seeliger, A. and Mühlhäuser, M. (2018a) 'Analyzing business process anomalies using autoencoders', *Machine Learning*, Vol. 107, No. 11, pp.1875–1893.
- Nolle, T., Seeliger, A. et al. (2018b) 'BiNet: multivariate business process anomaly detection using deep learning', in *16th ICBPM*, pp.271–287.
- Nour, M., Arabaci, B., Öcal, H. and Polat, K. (2024) 'New approaches to epileptic seizure prediction based on EEG signals using hybrid CNNs', *International Journal of Intelligent Engineering Informatics*, Vol. 12, No. 1, pp.85–102.
- Oketunji, A.F. (n.d.) *Evaluating the Efficiency of Hybrid Deep Learning Models in Distinguishing AI-Generated Text*, arXiv preprint arXiv:2311.15565.
- Pranata et al. (2023) 'Business relationship in business process management and management with the literature review method', *ADI JRI*, Vol. 5, No. 1Sp, pp.45–53.
- Roy, P. and Tamang, S. (2021) 'COVID-19 drugs invention using deep neural network models: an artificial intelligence approach', *International Journal of Intelligent Engineering Informatics*, Vol. 9, No. 2, pp.176–192.
- Saleheen, F., Habib, M.M. and Sejdini, A. (2024) 'Supply chain collaboration during disruption through sales & operations planning: a study in a manufacturing industry', *International Supply Chain Technology Journal*, Vol. 10, No. 1, p.332.
- Sarno, R., Sinaga, F. et al. (2020) 'Anomaly detection in business processes using process mining and fuzzy association rule learning', *JBD*, Vol. 7, No. 1, pp.1–19.
- Shamshiri, A., Ryu, K.R. and Park, J.Y. (2024) 'Text mining and natural language processing in construction', *Automation in Construction*, Vol. 158, No. 1, p.105200.
- Shreyashree, S., Sunagar, P., Rajarajeswari, S. and Kanavalli, A. (2022) 'A literature review on bidirectional encoder representations from transformers', *Inventive Computation and Information Technologies: Proceedings of ICICIT 2021*, pp.305–320.
- Smith, B. (2015) *Beginning JSON*, Apress [online] <http://books.google.com>.
- Taymouri et al. (2020) 'Predictive business process monitoring via generative adversarial nets: the case of next event prediction', in *18th ICBPM 2020*, pp.237–256.
- Tenschert, J. et al. (2020) 'Enterprise system capabilities for organizational change in the BPM life cycle', in *BPM Workshops*, Springer, pp.70–82.
- Tey, S. (2023) *Shooketh* [online] <https://github.com/steven-tey/shooketh?tab=readme-ov-file>.
- Townsend, J.T. (1971) 'Theoretical analysis of an alphabetic confusion matrix', *Perception & Psychophysics*, Vol. 9, No. 1, pp.40–50.
- van Dongen (2019) *Dataset BPI Challenge 2019.4tu.Centre for Research Data* [online] <https://doi.org/10.4121/uuid:d06aff4b-79f0-45e6-8ec8-e19730c248f1>.
- van Dongen et al. (2021) 'Conformance checking of mixed-paradigm process models', *IS*, Vol. 102.
- van Dun, C., Moder, L. et al. (2023) 'ProcessGAN: supporting the creation of business process improvement ideas through generative machine learning', *DSS*, Vol. 165.
- Varoquaux, G. and Colliot, O. (2023) 'Evaluating machine learning models and their diagnostic value', *Machine Learning for Brain Disorders*, pp.601–630.
- Zhang, H., Cheah, Y-N., Alyasiri, O.M. and An, J. (2024) 'Exploring aspect-based sentiment quadruple extraction with implicit aspects, opinions, and ChatGPT: a comprehensive survey', *Artificial Intelligence Review*, Vol. 57, No. 2, p.17.