INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Exploring the Ethical Accountability in Artificial Intelligence Negotiation Agents

Leonor Patrício Mendes

Master in International Management

Supervisor:

PhD, Pedro Miguel Ribeiro de Almeida Fontes Falcão,
Associate Professor with Habilitation
Iscte-Iul

November, 2024

BUSINESS
SCHOOL

Department of Marketing, Strategy and Operations;

Exploring the Ethical Accountability in Artificial Intelligence Negotiation Agents

Leonor Patrício Mendes

Master in International Management

Supervisor:

PhD, Pedro Miguel Ribeiro de Almeida Fontes Falcão,
Associate Professor with Habilitation
Iscte-Iul

November, 2024

Acknowledgements

During my academic journey writing this dissertation has been the biggest challenge. I have come to understand that this is not just a scientific exercise, it trains your organizational skills and your personal growth. However, after this long journey, I am proud of the work done and I hope it is useful for the academia.

To my family and to my partner, who have helped me get up on every challenge that this journey brought. A special thank you to my mother, who never gives up on me and does everything in her possession to see me smile.

To Professor Pedro Falcão who supported and supervised the work of this dissertation in all of its high and low moments. I'm very grateful for all the conversations and time you dedicated to this research.

To the eleven people I interviewed for this purpose. The availability of these was crucial and something to honour, since some conversations were more than half an hour. All the information given by each person helped me to draw conclusions and without it this study wouldn't go forward.

(page purposively left blank)

Resumo

A expansão da Inteligência Artificial (IA) para diversas áreas tem sido o cenário dos últimos anos. O campo da negociação não foi exceção, e com a integração de agentes de negociação baseados em IA, os seus processos tornaram-se mais eficientes, o que também levantou desafios éticos, especialmente em termos de responsabilização. Assim, a presente dissertação estuda a lacuna na compreensão de quem detém a responsabilidade, no ciclo de vida de um Agente de Negociação com IA. Esta é apresentada através de uma síntese de teoria da negociação, ética em IA e gestão de stakeholders, culminando no desenvolvimento de um quadro conceptual que mapeia papéis e responsabilidades, enquanto avalia, até que ponto os próprios Agentes de Negociação com IA, podem ser considerados responsáveis.

Através de uma análise qualitativa, incluindo entrevistas semiestruturadas com onze profissionais da indústria e académicos, foram identificados diferentes stakeholders chave como, por exemplo, Gestores de Projeto, Auditores Externos e Encarregados de Proteção de Dados. Categorizou-se igualmente os seus níveis de responsabilidade (Baixo, Médio, Alto) nas três fases principais do ciclo de vida do Agente de Negociação com IA: desenvolvimento, utilização e consequências. Adicionalmente, a pesquisa sugere que a responsabilização permanece centrada no ser humano, cabendo à Gestão Executiva, aos Gestores de Projeto e equipas técnicas, como a Equipa do Projeto, a primeira linha de responsabilidade pelos resultados do sistema. Contudo, os Utilizadores Finais partilham uma cota parte dessa responsabilidade, através da qualidade dos seus inputs e do consentimento sobre regulamentações. Por fim, os resultados proporcionam uma ponte prática para colmatar a lacuna entre complexidades técnicas e insights de gestão, para a tomada de decisão, sublinhando ainda os desafios contínuos nos sistemas de IA, como a fragmentação regulatória, falhas na governança de dados e a confiança do Utilizador Final nos resultados.

Palavras-chave: Agentes de Negociação com IA, Responsabilização, Responsabilidade, Stakeholders, IA Ética.

Códigos JEL:

- M10 Business Administration: General;
- M15 IT Management;

(page purposively left blank)

Abstract

The expansion of Artificial Intelligence (AI) to several areas has been the recent years setting. The negotiation field was not left out and with the integration of negotiation agents its processes became more efficient, which raised ethical challenges, especially in terms of accountability. Therefore, the present dissertation studies the gap in understanding who holds accountability in the lifecycle of an AI Negotiation Agent. It is presented through the synthesis of negotiation theory, AI ethics and stakeholder management, which culminated in the development of a conceptual framework that maps roles and responsibilities, while assessing the extent to which AI Negotiation Agents themselves can be held accountable.

By conducting qualitative research, including semi-structured interviews with eleven industry and academic professionals, key stakeholders were recognized, including Project Managers, External Auditors, and Data Protection Officers, and further categorized in their respective responsibility level (Low, Medium, High), across the three major phases of the AI Negotiation Agent lifecycle, development, usage and consequences. Furthermore, the research suggests that accountability resides on staying human-centric, where the Executive Management, Project Managers and the technical teams, such as the Project Team, bear the first layer of responsibility for the outcomes of the system. However, End Users detain a share on accountability by the quality of their prompts and consent over regulations. At last, the findings provide a practical bridge to fill the gap between the technical complexities and the managerial insights for decision-making, and it underlines the continuous challenges on AI systems, such as regulatory fragmentation, data governance breach and the End User's trust on the outcome.

Keywords: AI Negotiation Agents, Accountability, Responsibility, Stakeholders, Ethical AI

JEL Classification:

- M10 Business Administration: General;
- M15 IT Management;

(page purposively left blank)

Index

(page purposively left blank)

Figure Index

(page purposively left blank)

Table Index

(page purposively left blank)

Abbreviations List

AI – Artificial Intelligence

AI Act – European AI Regulation

ANAC – Automated Negotiating Agents Competition

API – Application Programming Interface

CRISP-DM – Cross-Industry Standard Process for Data Mining

DFNN / DFNNs – Deep Feedforward Neural Networks

DL – Deep Learning

DPO – Data Protection Officer

EU – European Union

GA – Genetic Algorithm

GDPR – General Data Protection Regulation

IT – Information Technology

ML – Machine Learning

MLPs – Multi-Layer Perceptrons

NN / NNs – Neural Networks

NLP – Natural Language Processing

NSS – Negotiation Support System

OECD – Organisation for Economic Co-operation and Development

PII – Personally Identifiable Information

PMBOK – Project Management Body of Knowledge

QA – Quality Assurance

RNN / RNNs – Recurrent Neural Networks

SOW – Statement of Work

UI – User Interface

US – United States

UX – User Experience

(page purposively left blank)

CHAPTER 1

Introduction

Negotiation has existed for centuries, and it can be described as the critical process to resolve conflicts and possibly achieve agreements by recognizing welfares and developing alternatives (Pienaar & Spoelstra, 1996). It is now imperative to human interaction, and it is profoundly studied by scholars on several views, such as within the organizational behaviour. Consequently, Dommen et al. (1966) cleared the understanding behind negotiation strategies, which included an explanation on the distributive tactic, where there is a maximization of the individual value (Weingart et al., 1990), and on the integrative approaches, where the cooperation allows to promote mutual gains (Park et al., 2019a; Pienaar & Spoelstra, 1996). Independently of the strategy chosen, there are a couple of steps that a negotiation should go through, the negotiation phases. These may be translated with different designations, although the version considered was proposed by Jonker et al. (2012), where a negotiation drives through preparation, bidding and closing. Additionally, the dynamics of each phase can be influenced by culture and trust, since these can impact the strategies used by both counterparts when sharing information for a better outcome (W. L. Adair et al., 2007; Rousseau et al., 1998). Hence, the present research has a highlight on the mentioned dynamics for a more effective negotiation.

Currently, Artificial Intelligence (AI) has become the buzzword for its transformative force in our society. It is no longer just an assistant on the minor tasks, these systems are shaping industries, are being designed to save lives and to make decisions for humans. Progressively, it is expanding to several domains and, one of them, is its integration onto negotiation processes in the business realm.

Consequently, research has codified negotiation practices, such as preference representation, giving origin to the intelligent negotiation agent (Aydoğan et al., 2020). AI Negotiation Agents have been designed to represent individuals or organizations by offering an automated negotiation that is efficient and adaptive to reach agreement (Lopes et al., 2008). Subsequently, their design integrates advanced algorithms that can handle dynamic and interactive preference specifications, allowing to enhance negotiation outcomes with a human counterpart (Kloe et al., 2020).

To comprehend this design, primarily it is essential to review the lifecycle of an AI project. It usually includes a few phases such as planning, data collection, algorithm development, deployment and monitoring (OECD, 2019; L. Yang et al., 2024). However, Miller (2022) has trimmed these phases into three major, development, usage, and consequences. According to the author, the development phase is centred on the system design, data processing, and algorithm training, possibly ensuring accuracy and biases mitigation (Martin, 2019; Miller, 2022; OECD, 2019). Subsequently, there is the usage phase, where the AI project relies on the End Users prompts to generate outputs, although there should be an ongoing monitoring of the system to guarantee adaptability and relevance over time (Vesa & Tienari, 2022). At last, during the consequences phase, the ethical implications, such as accountability and fairness, are reviewed to align with the societal requirements for a trustworthy AI system (Ryan & Stahl, 2021).

Additionally, the collective of these phases may frame an AI Negotiation Agent, which integrates multi-strategy selection, learning and language models (Bakhtin et al., 2022), showcasing a topic for further research.

In fact, the AI Negotiation Agent carries multiple benefits such as ending with constraints, like stress and inefficiency, associated with human negotiation (Baarslag et al., 2017a; Vij et al., 2015; Y. Yang et al., 2009). However, the growing deployment of AI Negotiation Agents has made the scientific body reflect on crucial matters about responsibility and accountability during its lifecycle. These are not traditional mediators during human negations, in truth, these AI systems are autonomous and can operate in ways that users find it difficult to comprehend, for instance, regarding the decision-making processes (Baarslag et al., 2017a).

Therefore, the ethical apprehensions are raised when mistakes, such as system failures leading to biases cases and ethical violations, occur (Falcão Filho, 2024). Addressing these challenges introduces a struggle to isolate responsibility and, consequently, the question of accountability, a “moral issue is present where a person’s actions, when freely performed, may harm or benefit others… A moral agent is a person who makes a moral decision, even though they may not recognize that moral issues are at stake” (Jones, 1991, p. 367).

Taking this into consideration, there is an essential need to review the stakeholders involved in the lifecycle of an AI Negotiation Agent, so it is possible to better comprehend if the agent itself is accountable. The research of Miller (2022) marks the start of a clearer definition of the wide range of roles and professions participating in the development of AI systems. However, the current body of literature is missing on a comprehensive framework, that provides an overview of the responsibilities of these stakeholders, throughout the AI lifecycle, especially when referring to negotiation scenarios. Finally, by attending the mentioned challenges, it could lead to a concrete and clearer path for responsible AI negotiation systems (Baarslag et al., 2017a; Floridi et al., 2018).

The present research addresses the mentioned gaps by giving an answer to the following questions: How can a conceptual framework be developed to effectively illustrate stakeholder roles and responsibilities across the AI Negotiation Agent lifecycle for managerial understanding? How far can AI Negotiation Agents be accountable for their actions? Therefore, it focuses on developing a conceptual framework, that can offer to the managers a, richer and more practical, comprehension on the dynamics between stakeholders and the lifecycle of an AI Negotiation Agent project. By achieving such, it allowed to connect the disparity between technical difficulties and the managerial decision-making requirements. The second goal is to understand if during the consequences stage the AI Negotiation Agent detains any level of accountability on its outcomes.

To accomplish the mentioned topics, it is necessary to study the subsequent objectives, where the first three are design to provide information for the conceptual framework, and the two left should complete the research on accountability:

- Refine the Conceptual Framework based on insights.
- Analyse the impact of key stakeholders during the lifecycle of the AI Negotiation Agent.
- Assess the mechanism needed for an improved involvement of the key stakeholders.
- Analyse the level of accountability of key stakeholders involved in the lifecycle of the AI Negotiation Agent.
- Analyse the impact of time on the level of accountability of the key stakeholders in the lifecycle of the AI Negotiation Agent.

The composition of the thesis is prepared so the reader can initially understand the main subjects in extant literature. The last carries authors which are on the working field, focusing their research on Negotiation, Artificial Intelligence Systems in Negotiation, including the AI Lifecycle with an emphasis on the Design and Development Phase, Ethics and Responsibility, and, at last, Stakeholder roles in AI projects. During this chapter it is essential to comprehend the characteristics behind an AI Negotiation Agent, from the negotiation strategies that it may use, to the algorithms chosen for its development. Furthermore, a first conceptual framework is proposed together with an explanation about the methodological approach taken. Subsequently, with the data collected it was possible to draft and compose the findings, by discussing them through the theoretical lens. At the last chapter, the main conclusions from the results are presented along with its limitations and with suggestions for future research.

CHAPTER 2

Literature Review

2.1. Negotiation

Negotiation is a process in which parties interact with each other in an effort to resolve conflict, despite profoundly divergent viewpoints, and reach a lasting agreement that is based on shared interests. It can be accomplished through the formation of common ground and the development of alternatives, commonly referred has people coming together, not merely through what they share mutually, according to Pienaar & Spoelstra (1996). Negotiation is a necessary part of life, just as we cannot communicate without it, we are hardly capable of functioning without it. Therefore, over the past fifteen years, organizational behaviour and management science have both placed a high priority on studying negotiation (Borbély & Caputo, 2017; Dommen et al., 1966).

In the words of Dommen et al. (1966), negotiation can be defined in two different strategies, the distributive strategy, and the integrative strategy. The first concerns the demanding of value as much as possible for the negotiator by influencing the counterpart on composing allowances (Mertes et al., 2023; Weingart et al., 1990). In contrast, integrative strategies denote not just the "win-win" scenario that is typically discussed (Pienaar & Spoelstra, 1996; Steinel & Harinck, 2020), but rather a "win more-win more" paradigm of negotiating. Thus, the goal for both sides is to leave with a sense of having acquired more than they could have with a different strategy. Accordingly, it is believed that conflicts are more expensive than reaching a compromise, as well as gains and losses should be balanced (Pinkley & Northcraft, 1994; Sticher, 2021). The expanding pie model, which states that both sides can have more with the same percentage share, if they cooperate to build a bigger pie, is a popular illustration of integrative bargaining (Park et al., 2019). Research after research, throughout cultures, has demonstrated that this collection of strategic behaviours is the most straightforward path to shared benefits and, therefore, most frequently used (Brett, 2002; Elms, 2006; Rinehart & Page, 1992). Consequently, it will also be considered as the negotiation strategy applied throughout this research.

Negotiation goes through different stages, and these can vary in the literature, however, for this purpose and according to Jonker et al. (2012), four phases will be considered: private preparation, joint exploration, bidding, and closing.

Private preparation is a segment where the negotiator strives to understand as much as possible about the opponent and the upcoming process, as well as its profile, and the negotiating area, meaning all the topics under discussion and concealed motives (Jonker et al., 2012; Peterson & Lucas, 2001). Following, the joint exploration phase, the parties engaged in negotiations communicate with one another but do not submit offers. The purposes of this stage are to verify the data they have already collected, provide a positive environment for the subsequent auction, expand the negotiating area, and decide on a bidding procedure (Lacher, 1981).

For the last practice, a negotiator should first invest in a bidding plan, to decide on the next offer and whether to accept an incoming proposal. Subsequently, if no further improvements can be expected, make a counteroffer, or stop, with the resource of an acceptance strategy. Finally, the closing stage can be reached, and the results of the bidding process are formalized and verified by both parties (Jonker et al., 2012). The negotiation goes back to one of the earlier phases if confirmation proves to be unattainable (Lim & Murnighan, 1994).

Occasionally, it is advantageous to move into a post-negotiation phase to verify if the agreement reached may be strengthened by outlining the preferences of both sides (Jonker et al., 2012; Broekens et al., 2010). The respective research will emphasize the bidding procedure, regarding decision-making, and the closing phase by studying their outcomes.

Culture and trust could be considered strong variables that influence the negotiation design, and the offers displayed brought to the table (Brett, 2000; Kong & Yao, 2019). These not only affect human-human interaction but similarly, human-agent relations. Adair et al. (2004) and Brett (2002) emphasize that different national cultures employ dissimilar negotiation strategies, which can include distinctive norms, values, and beliefs of a group as well as the ideologies underlying the social, political, and economic structures that assemble social interaction within the group. Therefore, the interests and priorities of negotiators are also influenced by their societal surroundings (Brett, 2017; Weingart et al., 2007).

In contrast to negotiators in cultures where the distributive approach is normative, integrative strategy practitioners tend to produce larger shared gains (Aslani et al., 2016; Gunia et al., 2011; Lügger et al., 2015). One exception exists, as W. L. Adair et al. (2007) proposed, negotiators in Japan employ a distributive method to reach agreements on shared profits, perhaps providing information about interests and priorities through implicit information exchange, leading to a mutual gain. It confirms the factor mentioned previously, culture, which influences the negotiation strategy.

Nevertheless, it all can come to trust, where one is disposed to expose oneself to another individual (Rousseau et al., 1998). Researchers have demonstrated that investing in trust during the negotiation eases information contribution (e.g., Kong et al., 2014). According to (Gunia et al., 2011), putting forth and responding to questions is an integrative approach that provides the other party with a chance to exploit the situation, and, subsequently, trust. It has an impact on negotiating tactics, interactions, and results. The same research also introduces the concept that inquiring exposes a negotiator's information gaps, which encourages susceptibility (Gunia et al., 2011). Consequently, negotiators resort to offering proposals and attempting to exert influence when there is a lack of trust (Gunia et al., 2011; Yao et al., 2017).

2.2. Artificial Intelligence Based-Systems in Negotiation

In the digitally connected world of today, technologies are advancing, and Artificial Intelligence (AI) is one of the primary innovations of the fourth industrial revolution (Macdonell et al., 2021), with its potential to imitate human behaviour (Fuchs et al., 2023). Thus, it has started to support complex domains, including negotiation.

To automate this sophisticated field, research has concentrated on codifying the negotiation process, such as domain and preference representation, as well as designing intelligent negotiating agents (Kröhling et al., 2024). These have been represented by acceptance strategies, opponent models, and bidding strategies (Aydoğan et al., 2020). Additionally, Jennings et al. (2001) studied three main research areas that can be involved, being negotiation protocols, negotiation objects, and agent decision-making models. According to the analysis, the negotiation protocol establishes the structured communication framework for parties, outlining rules and procedures. The negotiation object defines the specific issue being negotiated, such as price. Finally, the agent decision-making refers to the strategic choices negotiators make to achieve their goals within this framework (Jennings et al., 2001).

Negotiations with resources to automated agents can occur in multiple ways: human-to-human interaction, between agents, or agent-to-human negotiation (Cao et al., 2015). For the first scenario, a computer-mediated negotiation support system (NSS) is used, by employing decision-support systems (Kersten & Lai, 2007). These prioritize not only computer-based communication but also decision-making that holds assistance to a human negotiator. Consequently, the autonomous agent does not exercise control or autonomy (Cao et al., 2015; Kersten & Lai, 2007).

As opposed, software agents that represent both sides of a negotiation process are referred to as agent-to-agent negotiators. The purpose of these interactions is to persuade other agents to follow a particular course of action, alter an initial plan of procedure, or reach a consensus on an established strategy (Bala et al., 2015; Cao et al., 2015; Mirzayi et al., 2022). For the past few years, automated negotiation has been receiving increasing attention as a way to coordinate the interactions of computational autonomous agents (Chen & Weiss, 2012). Conversely, the study has referred to the interactions in a buyer-seller or consumer-provider relationship as usually showing conflicting interests over potential cooperative agreements.

At last, agent-to-human negotiation involves a different design for an automated negotiation agent (Chen & Su, 2022; Phogat & Tsumura, 2024). These require a method for making decisions, that adjusts and adapts to the human behaviour of the other negotiation party (Bala et al., 2015; Doğru et al., 2025). For scholars to obtain a deeper understanding of agent-to-human negotiations, it is imperative to investigate the possible impacts of the negotiation assignment's complexity (Stevens et al., 2018), considering that those interactions demand much more sophisticated agents (Cao et al., 2015). Hence, for a better comprehension of their complexed mechanisms, it is essential to have a deeper look at its lifecycle.

2.3. Artificial Intelligence Based-Systems Lifecycle

To better understand the concept of the AI negotiation agent, it is essential to have an overview of the lifecycle in an AI project. OECD (2019) started to gather all the essential phases in an AI project and defined the following order: planning and design, data collection, model building and interpretation, verification and validation, deployment, and, finally, operation and monitoring.

Similarly, L. Yang et al. (2024) stated that the AI project has six key phases, which include, planning, data collection, algorithm development, deployment maintenance and archiving. Then, Miller (2022) developed a more abroad approach where it splits the AI project lifecycle in three major phases, being development, usage and consequences, where each has their own stages and processes, such as training data on the first part, Figure 1.

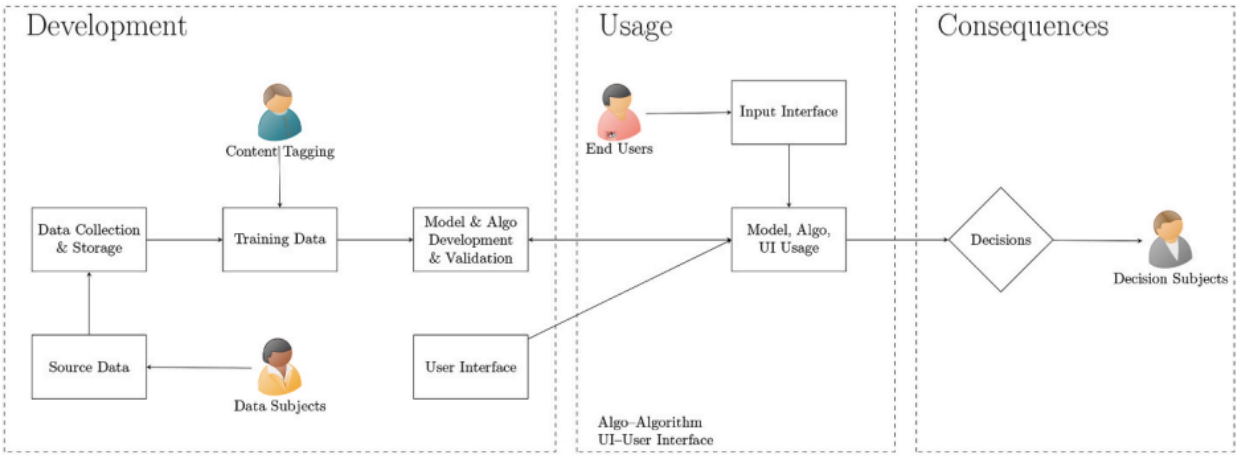

Figure 1 - AI development and usage project components. (Miller, 2022)

In this framework, the first major phase starts with planning and designing the system, and then it moves on to data collection and processing, along with sourcing it from, for instance, enterprise resource planning tools, websites and social media (OECD, 2019; Zarghami & Dumrak, 2021). Thus, these are large volumes of data that require human intervention, for labelling, tagging, or anonymization, as well as robust technologies to fulfil these processes (Bender et al., 2021; Miller, 2019; Moser et al., 2022). Afterwards, the subsets of this data, also known as training data, are used to build and validate the algorithms, with a mission to mitigate bias and improve accuracy (Martin, 2019; OECD, 2019). This work will later culminate on user interfaces enabling the interaction between the users and the algorithms on the format of, for example, chatbots, APIs, or smart devices (Simon, 2019; Vesa & Tienari, 2022).

Ending the development phase and entering the usage phase, end users translate as a crucial key actor, given that the AI system relies on them to place input parameters into the interface (Martin, 2019; Vesa & Tienari, 2022). These will then trigger the algorithm for outputs, even when the user is not aware of the operations behind it (Vesa & Tienari, 2022).

In this phase, there is also the process of continuous monitoring, where it is essential to detect issues and failures and to keep and improve the system performance. Additionally, time may play a role in the AI system, since, for instance, model values can become outdated or in need of a renewal process for constant relevance and functionality (OECD, 2019; Wieringa, 2020). Via this ongoing work, the organization can ensure that the AI system keeps adapting and evolving to the environment conditions.

Ultimately, Miller (2022) describes the consequences phase, where the decisions of the AI system may produce ethical considerations, such as in fairness, transparency and accountability that, in the end, evaluates its own outcomes (Jobin et al., 2019; Ryan & Stahl, 2021). This is the phase where there is an overview of how the outcomes can affect individuals, groups or even organizations, and how there is a need to ensure their alignment with societal and organizational requirements for a sustainable and trustworthy AI practice (Jobin et al., 2019; Ryan & Stahl, 2021).

2.4. Artificial Intelligence Based-Systems Design and Development Phase

Diving into depth on the development stage, may reinforce the comprehension on the difference between an AI project and an IT project of another kind. There isn't only the stages of data collection and training which require precision, but also the design of the model which can require a multifaceted approach, composed by multi-strategy selection, equilibrium strategies, learning techniques, and language models (Sun et al., 2024; Zhang et al., 2024). These elements, when synergistically integrated, empower agents to engage in effective negotiation with human counterparts.

The initial focus in exploring the components of the AI negotiation agent can lie on multi-strategy selection and adaptation, a foundational concept in agent-to-human negotiation (Li et al., 2021). Existing research highlights that these often employ a repertoire of strategies, such as time-dependent, behaviour-dependent, dynamic time-dependent, and impasse resolution strategies, demonstrating a superior performance compared to single-strategy models (Cao et al., 2020). For instance, in their experimental analysis, Haim et al. (2017), designed an AI negotiation agent adopting an equilibrium strategy in a three-player game that also involved human players.

As a result, the AI negotiation agents outperformed the human participants, where the topic of rational behaviours were considered in the strategy for both customer and service provider roles. Therefore, this multifaceted approach enhances the agent's effectiveness in negotiation, promoting higher settlement rates and optimizing joint outcomes for all parties involved (Cao et al., 2015; Li et al., 2021).

Furthermore, to anticipate the actions of their adversaries and, consequently, modifying their tactics appropriately, the AI negotiation agent can be reactive by integrating learning methods such as Genetic Algorithms (GAs) and Neural Networks (NNs) (Papaioannou et al., 2008; Rajavel & Thangarathanam, 2016).

The GA is used as an optimization method based on the process of natural selection, where the optimal or near-optimal solution is searched by mimicking the principles of evolution (Oreski & Androcec, 2020). It learns and perfects its concept of classification rules by interacting with the environment, resulting in a robust system architecture capable of dynamically adjusting its understanding of biases, making it adaptable to diverse concept categories (de Jong et al., 1993; Fernandez et al., 2010).

According to the research of Goldberg (2017), Neural Networks (NNs) belong to a powerful class of machine learning models, inspired on the human brain's structure and activity. The same author expressed that these can be divided in two major categories of Neural Network architectures: Feed-Forward Networks and Recurrent/Recursive Networks (Goldberg, 2017).

Feed-Forward Networks are identified as unidirectional information flows from input to output layers and it does not contemplate cycles or feedback loops (Sharma et al., 2013; Yalçın, 2021). The architecture consists of interconnected nodes, which are organized in layers with weighted connections encoding the network's knowledge (Sharma et al., 2013; Yalçın, 2021). For instance, the multi-layer perceptrons (MLPs) exhibit a notable capacity for processing both fixed-length and variable-length inputs, particularly in scenarios where the order of input elements is inconsequential and where exists datasets with varying input dimensions (Anonymous Authors, 2021; Goldberg, 2017). Through training on a given set of input components, MLPs acquire the ability to synthesize these components in a manner that facilitates the extraction of meaningful patterns (Goldberg, 2017).

In addition, there is a study suggesting that by adopting a Deep Feedforward Neural Networks (DFNNs), known as a hybrid strategy that combines machine learning algorithms with expert knowledge, the AI negotiation agent can outperform other models in predicting negotiation outcomes (Mell et al., 2021). This is possible considering that DFNNs are composed of fully connected layers of neurons with rectified linear unit activation functions and unbounded weights that can process simple operations and interact to build decisions (Lee et al., 2021; Liang et al., 2020).

However, when using DFNNs one should know that these produce models that carry hidden features and, when it comes to training data that may be biased, incomplete or influenced by past trials, the designers and developers may struggle to know and understand which parameters influence the model (Chasalow & Levy, 2021; Sambasivan et al., 2021).

On the other end, Recurrent Neural Networks (RNNs) possess directed cycles within their structure (Cardot, 2011; Grossberg, 2013). This unique characteristic endows RNNs with an inherent internal memory mechanism, making them particularly well-suited for processing sequential and time-series data (Cardot, 2011; DiPietro & Hager, 2020; Grossberg, 2013). According to Goldberg (2017), the RNNs functions as an input-transformer, undergoing training to generate informative representations for a subsequent Feed-Forward Network. This process effectively prepares the data for the Feed-Forward Network to try to predict values (Goldberg, 2017).

As stated earlier, some AI negotiations agents integrate Genetic Algorithms and Neural Networks to anticipate the actions of their adversaries. Upon revising each concept and for a better understanding of how both interconnects, Oreski & Androcec (2020) developed the following diagram, Figure 2. This approach automates the process of finding the best parameters for a Neural Network by employing a Genetic Algorithm, which systematically explores different parameter combinations to identify those that lead to the highest performance (Oreski & Androcec, 2020).

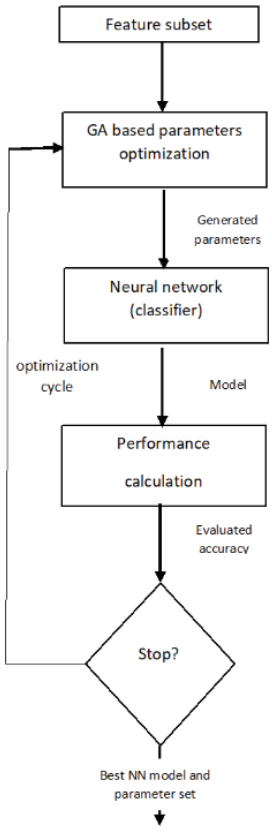

Figure 2 - Neural Network - Genetic Algorithm Architecture. *(Oreski & Androcec, 2020)*

Finally, it is crucial to integrate language models with strategic reasoning in AI agents for negotiations, so there is a natural communication with the human. The AI agent Cicero, for example, implements a language model in conjunction with planning and reinforcement learning algorithms to deduce the beliefs and intentions of participants and to produce dialogue as it pursues its objectives (Bakhtin et al., 2022). Given this integration, the agent may perform at a level comparable to that of a person in the cooperative and competitive strategy game (Bakhtin et al., 2022).

2.5. Opportunities and Challenges in Artificial Intelligence Negotiation Agents

The wide range of potential uses for automated negotiation in the various industries and disciplines is numerous and, consequently, the primary driver behind this emphasis (Chen & Weiss, 2012). Research has highlighted that this potential can translate in multiple advantages to the new type of market, the e-market (Park et al., 2019a).

For instance, automated negotiation can provide superior deals with a win-win agreement, by reducing in transaction and time expenses related to human action, as well as a growth in efficiency on the settlements that arise from the bargaining issues of semi-structured or multi-issue businesses (Chakraborty et al., 2020). Additionally, it is essential on diminishing few features that come with human negotiation, such as assisting negotiators in avoiding face-to-face meetings that could carry uncomfortable aspects, as in stress (Baarslag et al., 2017; Vij et al., 2015; Yang et al., 2009).

Furthermore, some studies even believe that autonomous negotiation will soon be a compulsory need, instead of just being preferred, considering that the human scale will not only lack rapidness but will also be costly (Mohammad & Nakadai, 2018). This exact scenario can be analysed on the renewable energy market and on the global placement of smart electrical grids, since these require dense and multiple contracts for each household (Baarslag et al., 2021).

Nevertheless, autonomous negotiators are still facing some challenges that tackle the effectiveness of these (Davidson et al., 2024). In their experimental analysis, Hunter (2015), concluded that an autonomous negotiator needs to interact with users to ensure that it builds an appropriate preference model to accurately represent it. Consequently, it prevents elicitation fatigue.

However, precaution should be used when communicating with users of negotiation systems, as they are frequently unwilling or unable to participate in them, leaving the model inaccurate (Baarslag et al., 2017). This is particularly crucial in areas involving privacy negotiations where users are noticeably reticent to interact with the system for extended periods (Baarslag et al., 2017). Therefore, future automated negotiators will need to, not only close transactions using the limited amount of user data that is now available, as additionally determine what more information should be requested from the user while causing the least amount of inconvenience to them (Baarslag & Kaisers, 2017; Memon et al., 2024).

Secondly, it can be challenging to translate from stochastic performance models of self-directed expert reasoning, into plain language that accurately communicates expectations and liabilities, since the negotiation agent's ability to reason may well surpass those of a no-specialist user (Ren et al., 2018). The referred can be translated into the acceptance of autonomous negotiators that has been limited by a natural conflict that is raised within autonomy and trust. Hence, the relevance of an autonomous negotiator is directly correlated with its capacity to act on its own, consequently influencing the user in meaningful ways (Yu et al., 2021).

Conversely, user confidence and disposition to cede control is commonly dependent on their comprehension of the agent's logic and the outcomes of its actions (Baarslag et al., 2017). The same study also suggests that the user can declare the acceptable means in the form of principles and increase the outcome space's transparency to bring both together.

Lastly, it is essential to acknowledge that people may bargain through mediators differently than they would in person. According to research on representation effects, when negotiating through a third party, a human one, people may behave less ethically and fairly (Chugh et al., 2005a). This begs the question of to what extent agents should equally mislead on behalf of a user, such as by employing argumentation and persuasive technologies (Burr et al., 2018; Dimopoulos & Moraitis, 2014).

Some might argue that negotiation agents should be moral, however, they should compromise these principles if it means maximizing their profits (Georgiadis et al., 2024). This argument is then analysed in a new context, outside the scope of negotiation, on ethical quandaries in self-driving automobiles.

The most effective approach to reconcile the inherent conflict, between accepting the agent's autonomy and accountability for its acts, is to accept user accountability for the agent's principles of design and targets (Baarslag et al., 2017). Thus, allowing to specify what should be accomplished and how. Consequently, the authors argue that this is another reason why humans should comprehend the agent, since taking responsibility for the agent's conduct indicates that one understands what the agent does (Baarslag et al., 2017).

Indeed, new studies on agent negotiators indicate that individuals might behave more effectively morally when bargaining through computer agents (de Melo et al., 2016; Duin & Pedersen, 2021). Nonetheless, additional investigation is required to comprehend the mechanisms underlying artificial representation effects.

2.6. Artificial Intelligence Ethics and Responsibility

Developing artificial intelligence with ethical frameworks in place has evolved into progressively more crucial as these machines grow in strength and integration into significant social systems. According to the statement by Picard & Picard (1997), ‘‘The greater the freedom of a machine, the more it will need moral standards.’’. Hence, some scholars contemplate that AI developers should carefully consider the ethics that are included in their systems, considering that the results of such scenarios may depend on their choices (Baum, 2020).

In fact, establishing moral decisions is believed to present the greatest obstacle for autonomous agents (Gill, 2020). The emerging discipline that examines the morality of machines, termed "Machine Ethics" or "Machine Morality" (Deng, 2015; Malle, 2016), has delved into issues such as how morality can be applied and to what extent autonomous machines should be moral beings.

Effectively, humans rarely make decisions that are consistent with how they claim to make them (Summerfield & Tsetsos, 2015). This is particularly so during negotiations when participants must choose a variety of actions for themselves and, later, these not only impact the company's success, reputation, and basic principles, but also serve as the foundation of their negotiation strategy (Mell et al., 2020).

Indeed, in real-world scenarios, people often reach inefficient agreements, even when the outcomes are physically efficient, as resolving conflicts can take significant time (Nakamura, 2024). When timing impacts utility levels, agreements may occur at inefficient points in the utility space, leading to a weak Pareto optimality (Nakamura, 2024). Furthermore, studies demonstrate that the mentioned strategy tends to be unethical, especially when people are given human intermediaries, such as lawyers, to act on their behalf (Chugh et al., 2005).

However, in a human-agent negotiation scenario, the AI agent is not designed to be a mediator, but to negotiate with people and propose several offers by applying a utility model. AI agents usually demonstrate an offer-proposal technique that complies with the majority of the principles and revised independence of irrelevant alternative property, which permits a range of potential solutions rather than a single, unique answer (Oshrat et al., 2009). Thus, the moral question is raised when an AI agent is programmed to negotiate on their behalf.

Malle (2016) and Gill (2020) suggested that moral cognition and affect, such as assigning condemnation and reasoning or moral decision-making, should all be coupled with robots' ethical competency. Similarly, de Melo et al. (2018) conducted a study in the Academy of Sciences by examining the norm of justice and demonstrating that when people interacted through AI systems treated each other more fairly. They discovered that when individuals "programmed" an agent to behave on their behalf across multiple domains, including negotiation, the agent acted more equitably towards others than they would have personally. This implies that individuals if given the choice, will choose to design AI agents who are models of moral behaviour, demonstrating a critical sensitivity to how their representatives perform (Mell et al., 2020).

Furthermore, a number of institutions have established a set of fundamental principles and regulations that should guide the creation and application of AI systems in society. Based on statements that produced a summary of 47 principles, Floridi et al. (2018) emphasized the similarities and significant variations of the concepts. Interestingly, all the arguments refer to the various approaches demonstrating an overlap and coherence to the four fundamental principles of bioethics: beneficence, non-maleficence, autonomy, and justice.

Additionally, as part of the European Commission's AI plan, the High-Level Expert Group on AI (2019) developed the Ethics Guidelines for Trustworthy AI, with recommendations that are built around three key assumptions: robust, ethical, and legal. In this very first regulation on AI, a framework that follows a risk-based approach was also designed, classifying an AI system with a potential risk level as: (1) unacceptable risk, (2) high risk, (3) limited risk and (4) minimal risk or no risk.

Consistent with such and to design AI systems that respect human autonomy, prevent harm, are fair, and are explicable, the mentioned ethical principles and associated values should be preserved.

However, it may be crucial to note that the same rules don't apply to the entire globe. For instance, the United States regulation on AI systems is based on a decentralized approach with initiatives that comply at a federal and state levels (Tovar Cardozo, 2023). Therefore, the AI Bill of Rights, doesn't follow the same principles as the European Commission's AI plan, leaving a gap between territories and, consequently, lacking at a unified framework (Castets-Renard, 2020; Tovar Cardozo, 2023).

Nevertheless, for the past years the regulations on AI systems have been evolving and adapting to the continuous innovation of these programs, with relevant efforts from the EU, US and China (Castets-Renard, 2020; Franks et al., 2024). In fact, very recently and after all the work already mentioned, the European Parliament has released their very first comprehensive AI law (Manchev, 2024). This regulation sets a pioneering regulatory standard aiming to ensure safety and legal compliance, aligned with the European Union fundamental rights and values. The law accentuates ethical AI development, compliance with copyright laws, transparency about training data, and robust cybersecurity measures, resulting in a possible balance of citizen rights protection with a welcoming competitive AI market (Manchev, 2024).

Despite some scholars agree that there are at least four core requirements, such as trustworthy, transparency, explainability and sustainable development, to obtain ethical systems (Ryan & Stahl, 2021), there are also a couple of challenges that arise from these.

Starting with the user's trust in the outcome, by following the incorporation of psychological factors into the utility function of the AI agent. The user requires an independent agent to provide a positive result, while the last one depends on the user for information and direction (Baarslag et al., 2017a). Consequently, the AI system can be seen as a black box, considering that, firstly, the interaction between the program and the end user can give results that may not be explained or interpreted by the last one (Cohen et al., 2014; Sambasivan et al., 2021). Secondly, the developers and the designers would be the ones to bridge this gap, although these may not understand as well some model decisions, leaving a lack of transparency and explainability (Sambasivan et al., 2021).

Nevertheless, trust in the system can be accomplished through co-participation, openness, and appropriate representation, resulting in reduced resistance to giving up control, and ensuring user satisfaction with and adherence to the intended outcome (Baarslag et al., 2017a). Additionally, (Okamura & Yamada, 2020) proposed the implementation of an adaptive trust calibration methods that may be useful to detect and mitigate improper trust levels. They argue that it can be accomplished by monitoring the user's behaviour and giving cues for trust recalibration, allowing the AI systems to prevent from over-reliance or under-utilization. Therefore, by starting to address these factors, negotiation settings may guarantee a more responsible AI-human interactions and greatly increase trust, mitigating this challenge.

Furthermore, responsibility and accountability for the actions of an AI system can come as a complex challenge. Has Wieringa (2020) stated on the 2020 Conference on Fairness, Accountability, and Transparency, AI are "systems capable of inflicting (minor to serious or even lethal) harms as well, be it intentional/unintentional". Therefore, determining who is accountable for the actions of AI systems can be complex. If an AI system makes a harmful decision, it can be unclear whether the responsibility lies with the developers, the users, or the AI itself. Establishing clear lines of accountability is an ongoing ethical challenge.

2.7. Stakeholder Roles in Artificial Intelligence Projects

The stakeholder word is frequently used on different fields of study and, specifically, in management. For a better understating of the concept, a few scholars have defined a stakeholder as "any group or individual who is affected by or can affect the achievement of an organization's objectives" (Freeman & McVea, 2001). Nevertheless, many theories have been erected to define who or what stakeholders are. For instance, the Project Management Body of Knowledge (PMBOK) sees the stakeholder as an entity associated with every project objective, meaning if they can help in it or harm it (PMI, 2021).

These can also perform roles, meaning that they behave in a context, guiding and directing it in a particular setting (Biddle, 1979; Zwikael & Meredith, 2018). Additionally, the categorization of stakeholder roles can be made in different models, however few take into consideration the passive stakeholder. This stakeholder category in a project serves to identify the ones that may be affected by the project's outcome, although they do not carry the power or opportunity to influence the outcome (Achterkamp & Vos, 2008; Habumuremyi & K Tarus, 2021).

Vos & Achterkamp (2006) stakeholder role-based model integrates the mentioned category and defines three roles, clients, decision-makers and designers, which belong to the active stakeholder category. According to the author, the active stakeholders are the ones who can affect the project and are involved, since there is a contribution of resources or influence in the outcome.

Furthermore, and taking this last framework into consideration, Miller (2022), studied who may be the stakeholders around an AI project. The author came with a representation of all possible AI community members, which were assigned to one of four stakeholder groups and six stakeholder project roles, as well as their interrelationships, Figure 3. In this research, every stakeholder was also categorized as active stakeholder, client, decision-makers and designers, as reference, as passive stakeholders and, lastly, as representatives, shown on Table 1.

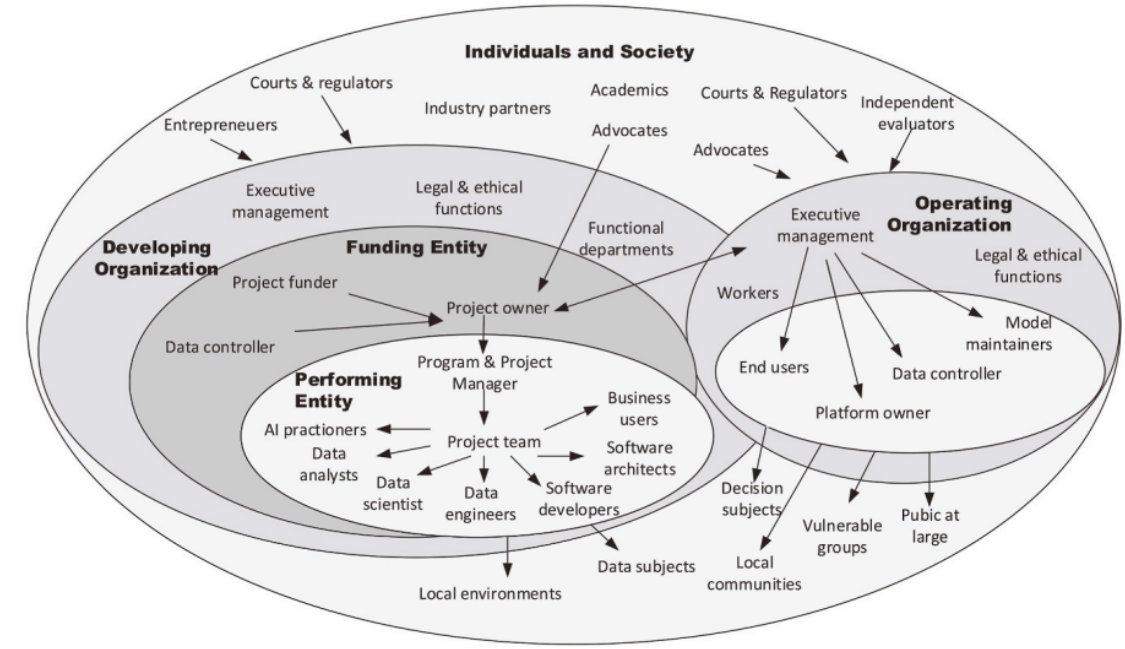

Figure 3 - The interrelationship between AI community members. (Miller, 2022)

Table 1 - Stakeholder project roles in AI project. (Miller, 2022)

Stakeholder project role	Project role
Client	Operating organization, model maintainers, platform owners
Reference	Data controller and data custodian, functional departments, legal and ethical functions
Decision maker	Developing organization, executive management, project funder project and program managers, project owner, project team, operating organization, *end users
Designer	Project owner, project team (business users, data analyst, data scientist, data engineer, software architects, software developers, AI practitioners)
Passive	Data subjects, decision subjects, *end users, entrepreneurs, local communities, local environments, public at large, vulnerable groups, workers
Representatives	Advocates, courts and regulators, independent evaluators, labor unions

Note: * Extremely context dependent.

Finally, Miller (2022) concluded that in AI projects teams make crucial decisions in building the artificial agents which can affect civil rights, humans and, consequently, life. Therefore, the author argues that ethical principles for fairness, accountability, transparency, and explainability in developing the AI project may not be enough and an inclusive stakeholder approach is needed.

CHAPTER 3

Research Methodology

While prior literature has underlined the design and stages of an AI program lifecycle, as well as meaningful challenges and advancements on ethical practices in these systems, there is still limited understanding of the concept of accountability in the mentioned domain. Miller (2022) researched who are the stakeholders in the AI community to understand their responsibilities and further accountability, although a gap persists, since the AI system can present an outcome which was not already predefined by their developers, unsettling the boundaries of responsibility within the stakeholders.

Therefore, addressing this gap may be important to improve the engagement and perception of accountability in the different roles of stakeholders. To do so, this study is directed by the following research questions:

RQ1: How can a conceptual framework be developed to effectively illustrate stakeholder roles and responsibilities across the AI Negotiation Agent lifecycle for managerial understanding?

RQ2: How far can AI negotiation agents be accountable for their actions?

According to the findings of Miller (2022) and the description of the AI project lifecycle that the author produced with resource to the OECD (2019) research, a diagram was developed, Figure 4. This diagram is an adaption from the work of the previous author, where all crucial stakeholders were gathered and placed in their respective stage of the AI project lifecycle. Therefore, there are three major stages of the AI project lifecycle: Development, Usage and Consequences. Lastly, Figure 4 was displayed for the following scenario: the Developing Organization conceives in-house AI Negotiation Agent to later be bought and implemented by the Operational Organization for their negotiation processes on a agent-human basis.

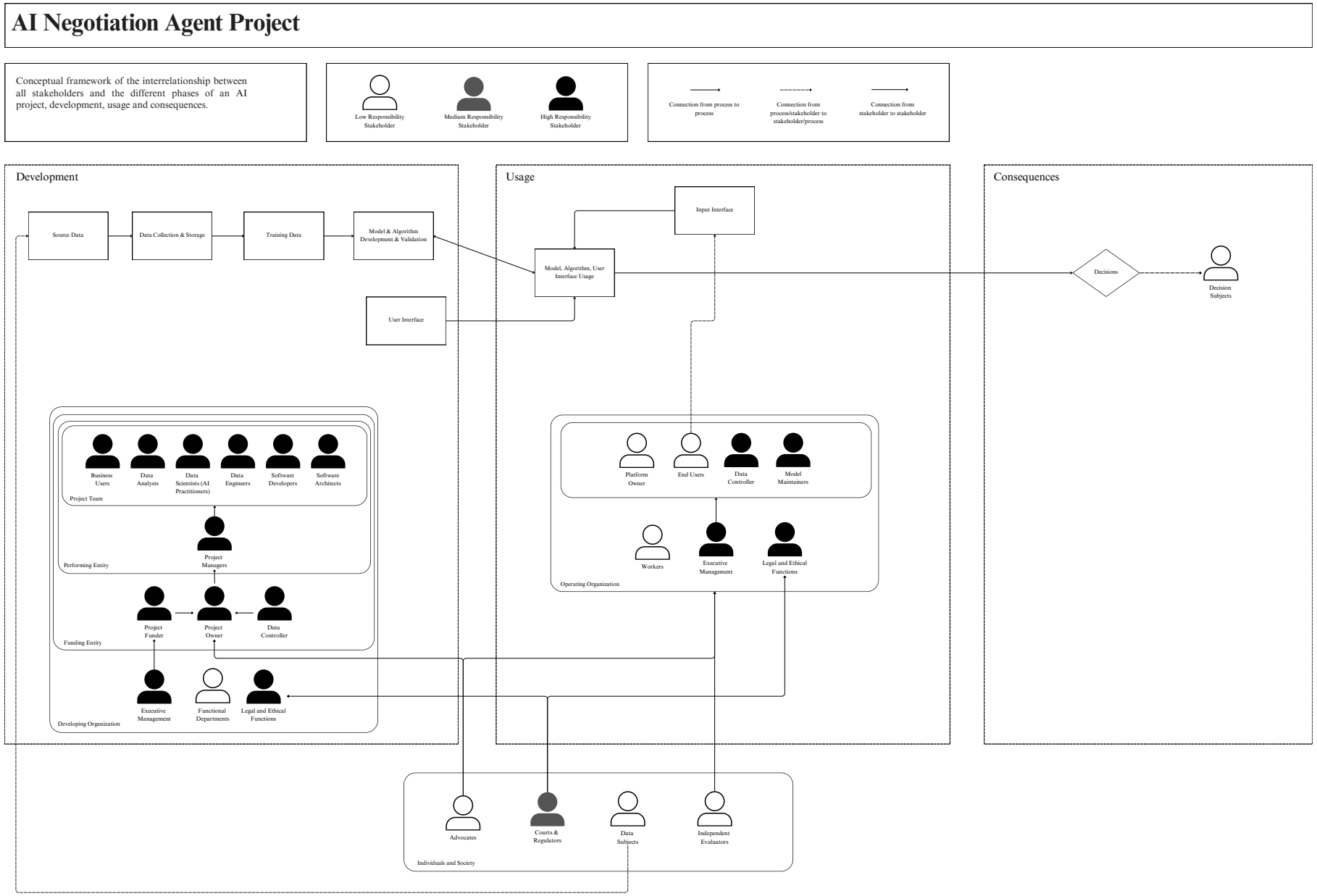

Figure 4 - Conceptual framework of the interrelationship between all stakeholders and the different phases of an AI project, development, usage and consequences. Authors own elaboration.

As to the human figures, each stakeholder has a colour regarding their level of responsibility. The colours translate into white, if a stakeholder was identified by Miller (2022) mainly has a passive stakeholder or Platform Owner, Independent Evaluators, Advocates or Functional Departments. Grey, if a stakeholder belongs to the Courts and Regulators category by the same author, and black, if a stakeholder is mainly an active stakeholder, belonging to the decision maker or designer groups from the mentioned study. Additionally, these levels of responsibility are classified as Low, Medium and High Responsibility.

The first one, white colour, englobes those who have little to no control over the project's direction or execution yet are indirectly touched by its results. These stakeholders play a more passive role and frequently depend on advocacy or representation to get their wants or concerns met.

Subsequently, in the medium level, grey colour, stakeholders don't make the strategic choices that determine the course of the project; therefore, their impact is substantial but limited. Frequently, their responsibility centres on making sure the AI project complies with outside legal, moral, or procedural requirements.

At last, in black colour, refers those who are involved in an AI project and who bear the greatest responsibility and influence over its success or failure. These stakeholders can control the project, and their choices may have a major effect on the outcome. The following table, Table 2, carries the aggregation of these stakeholders and their respective level of responsibility.

Furthermore, all the connections represented between stakeholders, and in between stakeholders and the different phases inside each AI lifecycle stage, were designed based on the work of Miller (2022). The definitions of each role in the project, provided by the author, are represented on Appendix A, which provides a better understanding of their placement.

Table 2 – Analysis of each Stakeholder considering their Stakeholder Category and Level of Responsibility based on the work of Miller (2022). Authors own elaboration.

Level of Responsibility	Stakeholder Category (Stakeholder project role)	Stakeholder (Project role)
Low	Client	Platform Owner
	Reference	Functional Departments
	Passive	Data Subjects
		Decision Subjects
		End Users
		Workers
	Representatives	Advocates
		Independent Evaluators
Medium	Representatives	Courts and Regulators
High	Client	Executive Management (operating organization)
		Legal and Ethical Functions (operating organization)
		Model Maintainers (operating organization)
		Data Controller (operating organization)
	Reference	Data controller (data custodian)
		Legal and Ethical Functions (developing organization)
	Decision Maker	Executive Management (developing organization)
		Project Funder
		Project Owner
		Project Manager
		Business Users (project team)
		Data Analysts (project team)
		Data Scientists/AI Practitioners (project team)
		Data Engineers (project team)
		Software Developers (project team)
		Software Architects (project team)
		Executive Management (operating organization)
		Legal and Ethical Functions (operating organization)
		Model Maintainers (operating organization)
		Data Controller (operating organization)
	Designer	Project Owner
		Project Team

After a careful revision of every topic on the literature review and the objective of this study, it was opted to pursue a qualitive research, where semi-structured interviews were chosen as the appropriate method for collecting information. This method allows the study to explore in depth and with flexibility the stakeholders involved in the decision-making and responsibilities across the AI program, Figure 5. It also allows to adapt the interview questions and flow based on new discoveries and information that arises during the interview, which translates in an essential approach for answering the Research Questions. In fact, the semi-structured interview, for some scholars, is considered to be one of the most powerful interviews for qualitative research, since it ensures that the one can develop subjective topics with more precision and evidence from the interviewees while keeping the focus of the study in consideration (Ruslin et al., 2022).

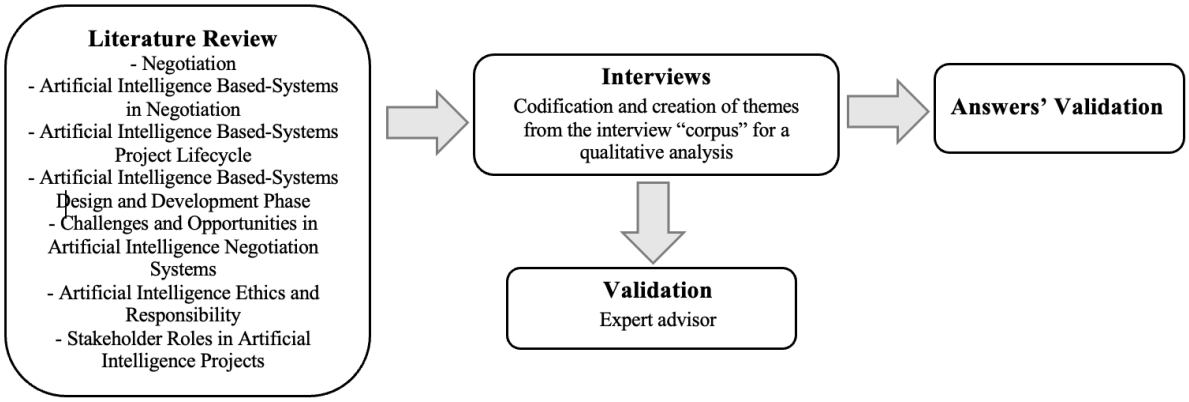

Figure 5 - Research and Model Design. Authors own elaboration

Consequently, it was made eleven interviews to individuals currently working on one of the following domains: Artificial Intelligence, Information Technology (IT) and IT Management. There were interviews, where the individual work or worked in more than just one domain of interest. Additionally, the interviews were conducted with the resource of a sample construction goal and the individuals were carefully picked, considering the mentioned subject. Nevertheless, there are many great answers, which for the current state are enough to produce conclusions.

The interviews were conducted during December-January 2024/2025, and they lasted from thirty minutes to one hour, see Table 3. These were done through video conference tools, such as Zoom and Google Meets, and the audio was recorded when allowed by the interviewee, to be further transcribed for analysis.

The informants were recruited with resource to the network of the researcher and additional contact suggestions provided by these. Permission to participate anonymously in the research was consented by the informants when scheduling the interview and at the beginning of the session. These are portrayed on Table 3 where the sample interviewees from an industry (n= 9) and academia (n= 2). Additionally, there were informants which had multiple roles during their career in IT, for example, a participant that has worked as software developer and is currently an IT consultant. Thus, the section "Time in this role" only refers to the last role this participant has performed.

Table 3 - Profiles of the interviewees. Authors own elaboration.

Interview Number	Age	Gender	Job Position	Time in this role (years)	Company Business	Company Industry	Date	Duration
1	26	Feminine	IT Senior Consultant	1,5	IT Consultancy	Service Provider	23/12/24	19:38:00
2	24	Feminine	Solution Engineer	2	IT Consultancy	Consultancy	27/12/24	41:16:00
3	25	Masculine	Information Systems Technician	3	Pension Payments	Banking	31/12/24	53:17:00
4	26	Feminine	Senior Solutions Engineer	3	Web Developer - Mobile System Integration	Consultancy	03/01/25	35:39:00
5	24	Masculine	Research Data Scientist	1	AI Fintech	Finance	06/01/25	33:43:00
6	50	Feminine	Associate Professor and Researcher	14	Education and Research	Education	07/01/25	41:50:00
7	35	Masculine	Full-Stack Developer	6	Master's degree Alliance	Education	08/01/25	50:05:00
8	31	Masculine	Automation Developer	2	IoT and Automation Consultancy	Consultancy	09/01/25	31:54:00
9	40	Masculine	Software Engineer	13	IoT and Automation Consultancy	Consultancy	09/01/25	59:34:00
10	38	Masculine	Automation and Control Systems Project Manager	11	IoT and Automation Consultancy	Consultancy	11/01/25	39:35:00
11	33	Masculine	Associate Professor and Researcher	8	Education and Research	Education	20/01/25	39:18:00

On the following table, Table 4, it is demonstrated the relation between the study objectives originated from the research questions and a summary of the main questions made to the informants with the purpose of fulfilling these aims.

Table 4 – Model which merges the objectives to answer the research questions and the respective interview questions. Authors own elaboration.

Research Questions	Objectives	Interview Questions
(RQ1) How can a conceptual framework be developed to effectively illustrate stakeholder roles and responsibilities across the AI Negotiation Agent lifecycle for managerial understanding?	(OBJ1) Refine the Conceptual Framework based on insights	(IQ1) Who are the stakeholders during the data collection and preparation process?
		(IQ3) Who do you consider to be the stakeholders during the planning and design process?
		(IQ6) Do you think there are roles for oversight or validation to ensure fairness and accountability during model development and testing? If yes, who assumes these roles?
		(IQ7) Who do you consider that ensures that the AI system meets organizational requirements during its implementation?
		(IQ10) Who is responsible for monitoring the AI system's performance and solving issues like bias, errors, or failures? For example, if a bias is detected, who do you think should handle it, and why?
	(OBJ2) Analyse the impact of key stakeholders during the lifecycle of the AI Negotiation Agent	(IQ2) What are the responsibilities of the stakeholders (e.g., data engineers, analysts) in ensuring data quality and privacy?
		(IQ4) What are their specific responsibilities during the planning and design process?
		(IQ5) What are the responsibilities of the technical team during model development and testing?
		(IQ8) What are the responsibilities of the Platform Owners or the Operational teams?
		(IQ12) What are the responsibilities of external stakeholders (e.g., regulators, lawyers) during the Consequences phase?
	(OBJ3) Assess the mechanism needed for an improved involvement of the key stakeholders	(IQ9) What mechanisms have you seen for collaboration among all stakeholders mentioned during the implementation and integration phase?
		(IQ13) How is feedback from external stakeholders, such as regulators or the public, incorporated?
		(IQ16) What would you suggest to improve the involvement of these 2 or 3 key stakeholders?
(RQ2) How far can AI negotiation agents be accountable for their actions?	(OBJ4) Analyse the level of accountability of key stakeholders involved in the lifecycle of the AI Negotiation Agent	(IQ10) Who is responsible for monitoring the AI system's performance and solving issues like bias, errors, or failures? For example, if a bias is detected, who do you think should handle it, and why?
		(IQ11) Is the End User responsible for the outcome of the AI system?
		(IQ15) If you had to summarize everything into 2 or 3 main stakeholders, who would they be?
	(OBJ5) Analyse the impact of time on the level of accountability of the key stakeholders in the lifecycle of the AI Negotiation Agent	(IQ14) Do you think accountability for the AI program's outcomes changes over time?

At last, a final question was made to the interviewees when the conversation allowed to do so. IQ17 is a more subjective interrogation, “Would you like to share any limits, challenges or interesting topics about AI systems?”, it is aimed to tackle matters outside of the stakeholder discussion. Additionally, to review the answers, a thematic analysis was chosen to identify codes, leading to categories and the global theme. Figure 6 displays the mentioned analysis where the theme was named “Challenges and Uncertainties in the Lifecycle of AI Systems” and it originated from three categories: “The Stakeholder’s Blind Spot”, “Constraints Across the AI Lifecycle”, and “Key Recommendations for an Ethical AI Systems”.

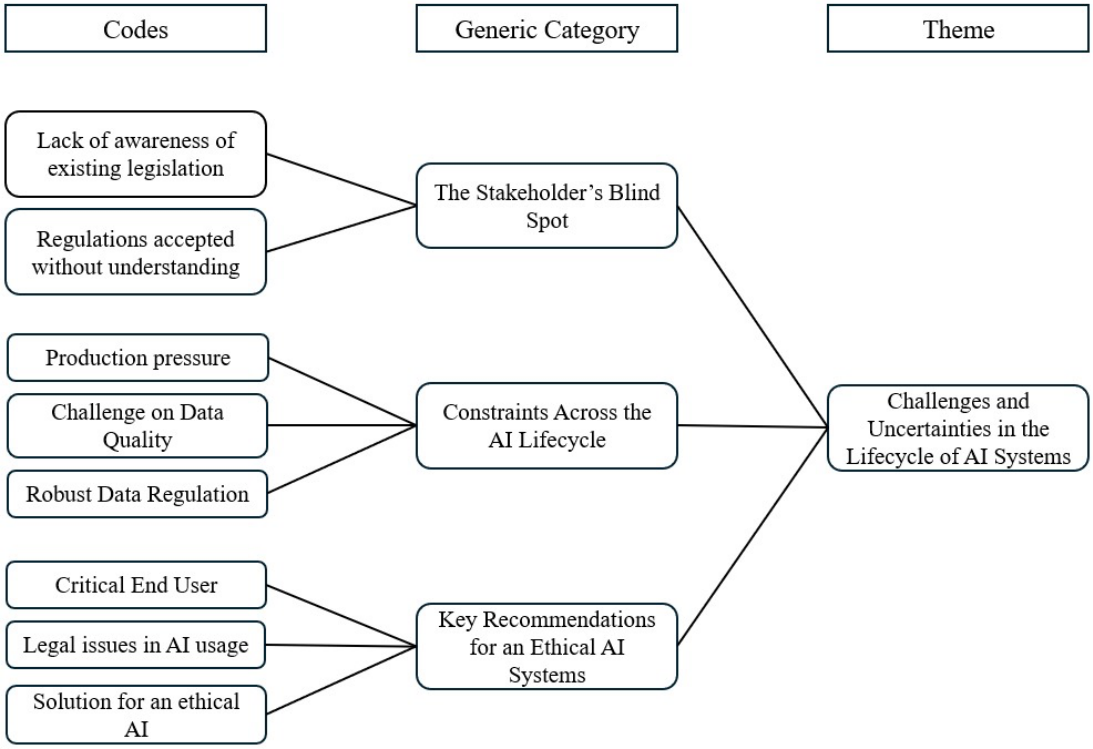

Figure 6 - Thematic analysis of the interview corpus. Authors own elaboration.

CHAPTER 4

Results and Discussion

4.1. How can a conceptual framework be developed to effectively illustrate stakeholder roles and responsibilities across the AI Negotiation Agent lifecycle for managerial understanding? (RQ1)

Figure 4, exposed previously, served as a beginning to unravel a discussion with the interviewees. Therefore, after a careful revision of the answers given, an improved version has arisen, Figure 7. One of the aims of this display is to provide a clear vision to managers, with a responsibility perspective, of the possible stakeholders and their interrelationships in an AI negotiation agent project. The lifecycle of this system inevitably has multiple technical aspects, which can be complex to comprehend and visualize. Thus, the mentioned diagram serves as a tool to the management team of an organization, by contributing on delivering a clearer comprehension of the impacts and consequences that an AI Negotiation Agent project can have during its lifecycle. Subsequently, to provide an answer to OBJ1 and OBJ2, the following changes were made:

Initially, ten out of the eleven interviewed mentioned the need to integrate the Business Understanding stage as the beginning of the lifecycle. As mentioned by interviewees six and eleven, on Appendix B.1, the six phases of CRISP-DM, Cross-Industry Standard Process for Data Mining, aren't contemplated in Figure 4. This methodology is an established method that can dramatically improve the outcome of the AI negotiation agent project (Hannecke, 2024).

Hence, it is necessary to, first, identify the business objectives by formulating the situation, as well as drafting all the data mining goals. Consequently, a document with all the requirements and plan should be produced for further guidance in the project. This modification can now be seen on Figure 7, where the Developing Organization was linked to the respective stage.

Secondly, during the stage of Data Understanding and Preparation, the key stakeholders identified by more than half of the informants where: Project Owner, Project Manager, Data Scientists, Data Engineers, Data Analysts, Software Architect and Data Subjects. A few answers, Appendix B.2, also pointed out to the Data Controller, Executive Management and Legal and Ethical Functions on the Developing Organization.

Additionally, a new stakeholder was mentioned, the Data Protector Officer, which was added to the diagram. In fact, it was presented as a crucial stakeholder since it ensures that all data processed, from Data Subjects or other source, complies with the applicable protection rules and no project should move forward without its permission.

Furthermore, in terms of responsibility, the Project Owner and the Project Manager were indicated as the ones with significant accountability during this process, “The owner of the project (…) has responsibility because they are the top (…) following, the project manager.” (Informant 2, Appendix B.2). On the topic of mechanisms used to guarantee data privacy and quality, the stakeholders mentioned previously from the Project Team potentially recur to the following (Appendix B.2): compliance with General Data Protection Regulation (GDPR) by encrypting data, user tests, traceable data and by avoiding data with Personal Identifiable Information (PII); anonymity contracts; and data pre-processing with, for example, filtering.

Following the Modelling stage where planning and designing occur before the development, the main stakeholders mentioned where the Project Owner, Project Manager and the entire Project Team, Appendix B.3. However, a new stakeholder emerged, the User Experience (UX) designer, since the AI Negotiation Agent has an interface to interact with humans. In this context, a new team has been added, the UX Team, composed by an UX Manager, UX researcher, UX and UI designers, information architect, usability expert, and UX writer. Nonetheless, this team may be operational after all the steps of the Modelling stage are completed.

Additionally, a different topic was discussed, the work methodology which will impact how this stage will run. Here one can have a waterfall methodology or, for instance, an agile methodology. This will dictate the pace of the project and the work of the Project Manager. Nevertheless, the one should guarantee that the deliverables are in place, report to the Project Owner and know every detail and process of the project. Three informants stated, on Appendix B.4, that the mentioned bit can be crucial for a successful flow of information between levels of hierarchy.

During the Evaluation period, more than half of the, on Appendix B.5, immediately pointed out a missing key stakeholder, the Quality Assurance Team (QA Team) composed by a QA Manager, Test Architect, QA Team Lead, QA Engineer and Test Analyst. A common flow in this phase would be described by three test levels, a first one within the Project Team, a second one done by the QA Team and a final one done by the client, on the scenario that the program was implemented outside of the Developing Organization.

By the end of this second level, the QA Team communicates to the Project Owner and Project Manager the results and these decide the steps that were validated and the ones that must go back to the Project Team for development. Therefore, on a higher level, these two last stakeholders are the ones that supervise the evaluation period. This flow can be traced out on Appendix B5 and B.6.

Concluding this period with success the project moves on to the Deployment stage. On the case that the future systems are to be used by an external organization, the Operational Organization, the question of who ensures that the organizational requirements are kept was raised (IQ7). The majority agreed that was again the responsibility of the Project Manager to guarantee that the first plan with all the requirements is set and ready for usage (Appendix B.7). Some argued that the Platform Owner, the Executive Management and the Legal and Ethical Functions on the Operational Organization side should also be involved in this process. These should assure that the system is working and keep a regular contact with the Developing Organization.

There are Operational Organizations that have a more technical team, the Model Maintainers, who can monitor for systems failures and consequently solve them, however, a commonly scenario is the non-existence of this stakeholder (Appendix B.8). Taking this into consideration, the informants mentioned the presence of a licensing contract, where the Project Team stays in charge of maintaining and monitoring the system for any kind of failures and of providing a continuous support to the Operational Organization (Appendix B.8).

At the Consequences phase, the discussion during the interviews demonstrates that the external stakeholders, even if they may not be directly involved during the development of the AI Negotiation Agent project, convey a considerable level of responsibility and further accountability, "(…) will have as much or more responsibility as any of the others I mentioned earlier (…)" (Informant 4, Appendix B.12).

Subsequently, two main external stakeholders were mentioned multiple times, Courts and Regulators and Independent Evaluators, specifically External Auditors (Appendix B.12 and B.13). The last does not exist independently on Figure 4, yet it was pointed out as fundamental to the AI project and it will be added separately from the Independent Evaluators. External Audits are in charge of assessing the project infrastructure, processes and if it complies with the existing standards and regulations. Therefore, they play a key role on identifying system failures and biased data.

In fact, in Europe, the new AI Act (Manchev, 2024) started to tackle the mandatory need to audit an AI project, “Regulated, now at the level of auditing, it must certainly be in the AI Act, detailing how those systems are audited.” (Informant 11, Appendix B.12), which can translate on the power and, consequently, the level of accountability of this stakeholder. Hence, Governments and Unions, such as the European Union, should also be considered for the research, since they can deliver regulation to create barriers and ensure the ethics behind AI.

Following the explanation of OBJ1 and OBJ2, OBJ3 comes as an approach to understand how the stakeholders and respective responsibilities are communicated. On an internal perspective and during the implementation stage, three major tools were discussed (Appendix B9). A communication platform, such as Teams or Zoom, a management platform, such as Trello or Monday, and a “Help Desk” platform, such as a ticket system. However, documenting all the steps and tests of the system was identified as essential, since it allows the higher levels of hierarchy to be updated. The mentioned topic can also be accomplished by committees, where all the evolutions, struggles and next steps of the project are presented to the Executive Management. Additionally, the mentioned mechanisms were cited once more when the discussion led to the involvement of the two or three key stakeholders (IQ16, Appendix B.16).

Nevertheless, during this conversation there was a challenge mentioned by four informants, the fact that information is commonly trimmed when it comes from the Project Team to the higher hierarchies, leaving space for losing some appealing details (Appendix B.16). In fact, even when using a ticket system internally, the informants weren’t satisfied with the flow of information. Therefore, the need to have a frequent communication arises and regular meetings between the Project Manager with the Project Owner, both with Executive Management and again both with the Operating Organization, such as the Platform Owner, could be perceived as a solution for this challenge.

In terms of an external perspective, it is key to analyse how does the information, especially feedback, from the external stakeholders is acknowledge by the Developing and Operational Organizations. The public opinion, according to the informant 3 on Appendix B.13, can be obtained by doing tests with a sample, or opinion surveys, or polls. It was also mentioned that an official website could be a source of information, since usually there is a section to leave feedback or contact the organization, which then can be led to a ticketing system internally.

However, when it comes to judicial communication the mechanisms are slightly different, for instance, "(…) there are the courts and regulators and if a public notice of opening is published (…)" (Informant 4, Appendix B.13). Consequently, when these kinds of external stakeholders use mechanisms to communicate, it comes in an imperative status and the need to assist it is eminent. Thus, for the majority informants, the judicial external stakeholders, such as Courts and Regulators or Independent Evaluators, carry a significant level of responsibility.

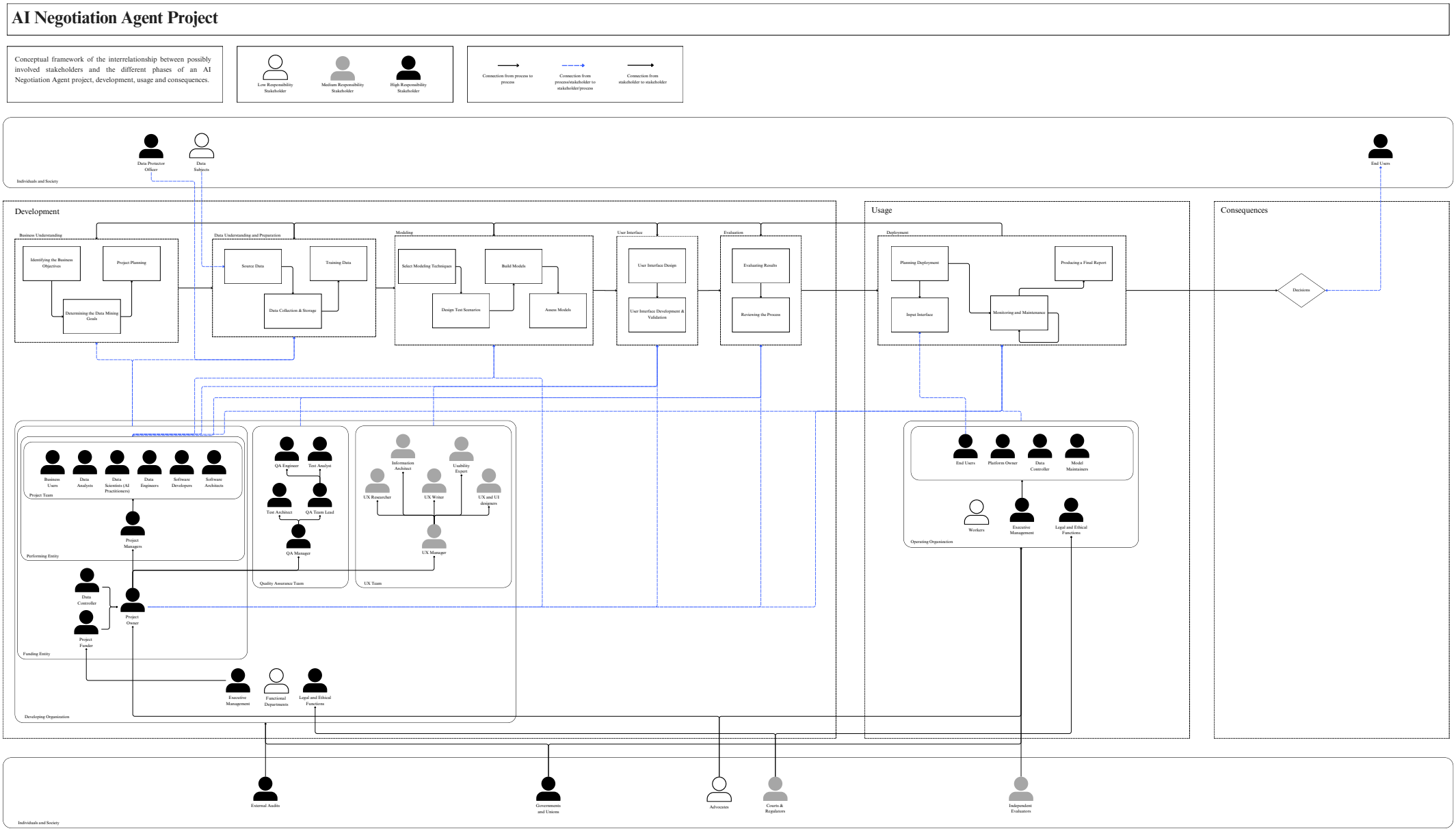

Figure 7 - Conceptual framework developed to effectively illustrate stakeholder roles and responsibilities across the AI Negotiation Agent lifecycle for managerial understanding. Authors own elaboration.

4.2. How far can AI negotiation agents be accountable for their actions? (RQ2)

In an effort to provide an answer to the RQ2, a first objective was designed, OBJ4, to analyse the stakeholder's accountability in one of the most critical scenarios, when the AI system fails. Additionally, understanding until where is the End User responsible could be an extra to the answer, as well as who would be pointed out when we try to summarize this considerable list of stakeholders, just into two or three. These questions were imposed considering that the AI system is already available to the public and in use, hence it is referring to the Consequences phase.

The Appendix B.10 illustrates the answers given by the informants to the mentioned question. There isn't a global consensus on who is accountable for handling the failures of the AI system. Some mentioned that the Project Team is the one accountable, where others consider the Platform Owner or even the entire Operating Organization. However, the overall agreed that there should be a maintenance team either on the Developing Organization or on the Operating Organization side.

This changes with the licensing contract previously stated, where there is a maintenance period which defines the organization responsible for the AI system. The Operating Organization may have a Model Maintainers team, which then stays responsible for the maintenance, or it may not be large enough to detain these structures and it requires the assistance of the Developing Organization or a third party. Nevertheless, the declared maintenance team is the responsible group of stakeholders for monitoring the system's performance, to solve issues or even to trigger a flag to the Project Team, "(...), you usually have a team responsible for actively evaluating the model in production or something similar, with methods to continuously assess or detect anomalies so they can then make corrections." (Informant 5, Appendix B.10).

Apart from the stage of monitoring and maintenance, there is another crucial stakeholder during the Consequences phase, the End User. Considering that this is not a simple program where the End User gives an input and there is already a programmed output for that specific prompt, the AI system carries other algorithms which allows a multi-strategy selection and an adaptation to the prompts. Therefore, does it mean that the End User may be responsible for the outcome of the AI system?

Eight out of eleven informants, agree so, “It seems to me that he is responsible because, well, it’s like a recipe, isn’t it? Depending on the ingredients you put in, it can have various outcomes.” (Informant 3, Appendix B.11). Thus, it states the initiative that the stakeholder holds the accountability, disregarding the condition of the outcome. Additionally, those who don’t consider that the End User is responsible, they raised a relevant topic (Appendix B.11). Frequently, there is a set of statements that the AI system displays and are agreed when the End User creates an account or begins the usage period. These may ensure that all the information given to the AI system is saved in a memory, even when the End User isn’t completely conscious about it. Therefore, when providing sensible information, the End User is permitting a leaking of data without its “conscious consent”. It may be interpreted as a lack of knowledge of the existing legislation.

Subsequently, on the framework of accountability, it was questioned the two or three key stakeholders during the whole lifecycle of an AI project and the most mentioned were the Executive Management, (n=6), the Project Manager (n=4) and the Project Team (n= 5), which can be seen on Appendix B15. Hence, these translate into the core stakeholders when it comes to accountability for the outcomes of an AI system. However, and considering the topic on licensing contracts, mentioned before, it was necessary to analyse the possibility of a variation on these stakeholders when it comes to time. The unanimity thinks that time doesn’t affect accountability and that the declared stakeholders are the ones responsible for the AI system throughout the years.

At last, no informant placed the accountability on the system itself, “These machines will not come to life on their own unless we want them to reinvent themselves, and for that, they must have some instruction to do so.” (Informant 11, Appendix B.17). Therefore, and replying to the RQ2, AI Negotiation Agents are not accountable for their outcomes, instead the Executive Management, from both Development and Operating Organizations, the Project Manager and the Project Team, are the ones liable for it.

4.3. Challenges and Uncertainties in the Lifecycle of AI Systems

The fast-paced innovation around AI has challenged the speed of legislation to follow the changes of this technology. In fact, nine out of eleven interviewees, on Appendix B.17, had a word about this topic. Multiple times during the conversation informants showed a lack of knowledge about the current regulation on Data and AI. These were confident that the rules exist, although they mentioned that it wasn’t clear on the quotidian of the Project Team, “Of

course, there should be something in place, right? Legally, there must be a way to determine who is truly at fault and to what extent. But I don't think this is something that's very present in the team's day-to-day work. I believe that, in their daily routine, the team always feels responsible." (Informant 2, Appendix B.17). However, there were notions about GDPR and PII's, but it is not deep enough, meaning that it could challenge the ethical and trustworthiness of an AI system at its core. Hence, a code originates from these words, "Lack of awareness of existing legislation".

These circumstances, to an extent, apply as well to the End Users when these accept the disclaimer or terms and conditions of the AI system, without fully comprehend the consequences of it, "People thought it was something traditional; they had no idea what was actually happening, and they kept inputting data." (Informant 9, Appendix B.17). This has been discussed previously on the analysis of the RQ2, however informant 11 (Appendix B.17) added a particularity, "Ignorance leads to poor judgment, and poor judgment leads to actions that can be harmful.", the power of the stakeholders that develop in comparison to the external which lack knowledge leaves a significant gap and possibly allowing manipulation.

During this interview, the "Black Box" concept was discussed in terms of a tool to scare and "control" the public opinion, rather than coming from lack of transparency by the AI system. Thus, a new code is dictated, "Regulations accepted without understanding". Consequently, both codes, "Lack of awareness of existing legislation" and "Regulations accepted without understanding" form a first category named "The Stakeholder's Blind Spot".

Furthermore, the informants which work on project teams, either as developers or engineers, stated the fact that in some scenarios there is a rush on producing and delivering, "The biggest pressure is time, so projects tend to be more complex, as short as possible, to deliver the product as quickly as possible." (Informant 9, Appendix B.17). The recurrency of such concedes place for mistakes and consequences, leading to the question of "Is the Project Team and the Project Manager comfortable with afterward consequences?". However, there may exist the intervention of an External Audit to correct these, although probabilities still play a role on how many system issues are detected and corrected. The code "Pression in production" arises along with the challenge during the development phase of the AI system.

On the search for system mistakes, informant 10 on Appendix B.17 revealed that for an AI system data quality is crucial for its operation, "(…) but yes, I detected those gaps; sometimes, a subtle change is enough, and artificial intelligence no longer has the competence to handle it.". Therefore, the Data Understanding and Monitoring stages cannot be rushed, if data lacks on quality the outcome may lead to severe consequences.

Subsequently, there is a concern on the informants about the continuous work on these systems, especially in terms of data, since this is not a computer program that is created and at no time reviewed, thus the code “Challenge on Data Quality”.

A second code referred to data was created, “Robust Data Regulation” to highlight statements such as “Who will store these data? (…) there could be a third party involved, (…), who has servers and will store all that information from the application. They end up being the data holders, and often they make backups of this data. Neither the client nor the developer, in this case, is responsible for it, and there could be a data leak or something else, causing this data to end up somewhere it shouldn't.” (Informant 9, Appendix B.17). The cited promotes the question about the journey of all the data used or produced in an AI system.

Despite the memory slots that the Developing Organization possesses, there is the need, particularly when referring to online systems, to access a server. The Developing Organization may not detain this server, resorting to an external third party, hence the existence of a robust data governance framework may be crucial. Finally, these three codes, “Pression in production”, “Challenge on Data Quality” and “Robust Data Regulation” originate the category “Constraints Across the AI Lifecycle”.

Ultimately, a final category was formed, “Key Recommendations for an Ethical AI Systems”, which englobes three codes. The first one, “Critical End User”, appeared from expressions such as “(…) we have to review what's it gave you (…).” (Informant 7, Appendix B.17) and “(…) if you’re reading an article, you believe it because it’s plausible, right?” (Informant 8, Appendix B.17), where the End User frequently assumes that all information given by the AI system may be true.

Consequently, these may require a deeper education on the usage of AI systems, and criticism on its outcomes is essential, translating as a recommendation for the years to come. Additionally, in terms of regulation, informant 6 on Appendix B.17 suggested that there should be a continuous update on the existent laws, since these systems are very complex, and they will continue to be so. Thus, the legal sector should actively track the curve of progression, and it may gradually turn into a challenge. To represent it, the code “Legal issues in AI usage” is created. The category is completed with the code “Solution for an ethical AI”, which represents a mechanism that may tackle the previous concerns about data quality and similar issues.

Lastly, the resource to software updates is a recommended process to amend any mistakes formerly done by the Developing Organization, according to informant 9 and 7 on Appendix B.17, “You'll remember that mobile phones (…). Computers (…). Now we even have cars with updates. So, unfortunately, this is the reality of software (…).” and “Have a review of the code to repair”.

(page left purposely blank)

CHAPTER 5

Conclusions

The negotiation environment has suffered an evolution with the introduction of artificial intelligence systems, such as AI negotiation agents. These systems have the power to inflict "minor to serious or even lethal harms as well, be it intentional/unintentional" (Wieringa, 2020), therefore the challenge behind accountability has become more complex. The present research explores the gaps on ethical allusions, responsibility and further accountability, associated with AI negotiation agents. The analysis of both attributes is done during the lifecycle of the system, meaning from the early stage of development until the usage and consequences phase. Inevitably, a study of the stakeholder roles was conducted, since these are entities related to the project objectives, even in terms of helping it or harming it (PMI, 2021). By completing this, it was possible to map every stakeholder that may be involved to understand the challenges surrounding AI accountability on its moral decisions (Gill, 2020) and additional ethical considerations behind the key roles, primarily introduced by the research of Miller (2022).

To further investigate the stated topics, a semi-structured interview was conducted to eleven interviewees that have worked, or are working, in one or multiple areas related to Artificial Intelligence, Information Technology (IT) and IT Management. These sessions lasted from thirty minutes do one hour and the conversation started with resource to a conceptual framework on the stakeholders' roles involved during the lifecycle of an AI project, which was previously developed from the work of Miller (2022).

There are three major key findings, which reinforce the attribution of responsibility concerns that were raised in the ethical AI literature (Ryan & Stahl, 2021; Wieringa, 2020). First, starting with a refined conceptual framework which maps new roles, such as Data Protection Officer, QA Teams and External Auditors, besides the ones formerly studied and additional technical stages of the lifecycle. This tool not only allows to visually understand the interrelations and responsibilities across the phases of an AI Negotiation Agent, has it highlighted the need for structured accountability mechanisms. Additionally, the framework bridges the complexity of the technical domain with the managerial field when it comes to decision-making, providing the necessary clarity on accountability across a multi-stakeholder project. Thus, it supports former research on stakeholder engagement in AI projects, emphasizing the need of supervision at the various phases of development and operation (Miller, 2022).

Furthermore, the second finding suggests that AI systems can indeed deliver efficient negotiation outcomes, however accountability persists on being human-centric. Nevertheless, responsibility is not attributed to only one stakeholder, such as developers, on the contrary, it extends across a range of stakeholders, compromising Executive Management, Project Managers and technical teams, such as the Project Team, on a primary level of responsibility for the outcomes. External stakeholders, such as End Users, Regulators and auditors share partial accountability, either through input quality and consent, or by law enforcement. This notion shares a similar assessment that accountability frameworks should integrate multiple levels of supervision structures studied by OECD (2019) and Manchev (2024).

Lastly, findings reveal that AI systems apart from their autonomous benefits, still raise challenging ethical uncertainties in terms of transparency and trustworthiness (Baarslag et al., 2017a; Floridi et al., 2018; Sambasivan et al., 2021). These barriers can be seen in scenarios such as regulatory fragmentation and data governance gaps, which underlines the continuous necessity for adaptive legal frameworks, strong audits and the education of the End User, to possibly mitigate risks affecting individuals and society.

The previous findings contribute to the academia and practice by extending the model developed by Miller (2022) with the addition of new roles, technical stages and a responsibility gradient that categorizes the accountability of each stakeholder. Supplementary and practical tools are provided by synthetizing mechanisms, such as agile collaboration and data encryption, that can be implemented by organizations when developing a more ethical AI Negotiation Agent. Additionally, the study combines negotiation theory with AI ethics and pushes "Machine Ethics" (Deng, 2015; Malle, 2016) by showing how the stakeholder involvement may give the solution for conflicts between autonomy and accountability.

Regardless of the mentioned contributions, the present study has its own limitations, starting with the sample size, since it is not vast. Nevertheless, it contains great answers from the interviewees, although it could go a step further and future studies may validate these findings with a wider participation. In addition, qualitative data gathered has its own pitfalls, such as it can easily be outdated, since the topic on AI is in constant evolution, and the culture aspect, considering that the majority interviewed were Portuguese and AI ethics are not studied the same across the globe. Therefore, with this gap in mind, it is suggested that future research addresses these topics with a mixed-method approach to possibly reach informants outside of Europe.

Concluding the topic on future research, it could be interesting to test the refined conceptual framework in real negotiation scenarios, such as in e-commerce, along with regulatory frameworks, such as the EU AI Act, to analyse the impact of AI Negotiation Agents in accountability on practice. Additionally, the comprehension of how are AI principles actually integrated on the pipeline of development, could be further explored (High-Level Expert Group on AI, 2019). Lastly, a more adaptive juridical approach on AI liability models would be crucial research to accelerate the creation and release of regulation promoting ethical systems. Moreover, by addressing these gaps the path to ensuring responsible AI Negotiation Agents is underlined and it drives to less harm for society and individuals.

(page left purposely blank)

References

Achterkamp, M. C., & Vos, J. F. J. (2008). Investigating the use of the stakeholder notion in project management literature, a meta-analysis. *International Journal of Project Management*, *26*(7), 749–757. https://doi.org/10.1016/j.ijproman.2007.10.001

Adair, W., Brett, J., Lempereur, A., Okumura, T., Shikhirev, P., Tinsley, C., & Lytle, A. (2004). Culture and Negotiation Strategy. *Negotiation Journal*, *20*(1), 87–111. https://doi.org/10.1111/j.1571-9979.2004.00008.x

Adair, W. L., Weingart, L., & Brett, J. (2007). The timing and function of offers in U.S. and Japanese negotiations. *Journal of Applied Psychology*, *92*(4), 1056–1068. https://doi.org/10.1037/0021-9010.92.4.1056

Anonymous Authors. (2021). *MLP-BASED ARCHITECTURE WITH VARIABLE LENGTH INPUT FOR AUTOMATIC SPEECH RECOGNITION*. https://consensus.app/papers/mlpbased-architecture-with-variable-length-input-for/117901b9848a58fea0486c5cd0e7b527/

Aslani, S., Ramirez-Marin, J., Brett, J., Yao, J., Semnani-Azad, Z., Zhang, Z., Tinsley, C., Weingart, L., & Adair, W. (2016). Dignity, face, and honor cultures: A study of negotiation strategy and outcomes in three cultures. *Journal of Organizational Behavior*, *37*(8), 1178–1201. https://doi.org/10.1002/job.2095

Aydoğan, R., Baarslag, T., Fujita, K., Mell, J., Gratch, J., de Jonge, D., Mohammad, Y., Nakadai, S., Morinaga, S., Osawa, H., Aranha, C., & Jonker, C. M. (2020). Challenges and Main Results of the Automated Negotiating Agents Competition (ANAC) 2019. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12520 LNAI*, 366–381. https://doi.org/10.1007/978-3-030-66412-1_23

Baarslag, T., & Kaisers, M. (2017). *The value of information in automated negotiation: a decision model for eliciting user preferences*.

Baarslag, T., Kaisers, M., Gerding, E. H., Jonker, C. M., & Gratch, J. (2017a). When Will Negotiation Agents Be Able to Represent Us? The Challenges and Opportunities for Autonomous Negotiators. *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, 4684–4690.

Baarslag, T., Kaisers, M., Gerding, E. H., Jonker, C. M., & Gratch, J. (2017b). When Will Negotiation Agents Be Able to Represent Us? The Challenges and Opportunities for Autonomous Negotiators. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, 4684–4690. https://doi.org/10.24963/ijcai.2017/653

Baarslag, T., Kaisers, M., Gerding, E., Jonker, C., & Gratch, J. (2021). *Self-sufficient, self-directed, and interdependent negotiation systems: a roadmap towards autonomous negotiation agents*. https://consensus.app/papers/selfsufficient-selfdirected-and-interdependent-baarslag-kaisers/dafcb57560a959aca21246728f77d7d5/

Bakhtin, A., Brown, N., Dinan, E., Farina, G., Flaherty, C., Fried, D., Goff, A., Gray, J., Hu, H., Jacob, A. P., Komeili, M., Konath, K., Kwon, M., Lerer, A., Lewis, M., Miller, A. H., Mitts, S., Renduchintala, A., Roller, S., … Zijlstra, M. (2022). Human-level play in the game of Diplomacy by combining language models with strategic reasoning. *Science*, *378*(6624), 1067–1074. https://doi.org/10.1126/science.ade9097

Bala, M. I., Vij, S., & Mukhopadhyay, D. (2015). Automated negotiation with behaviour prediction. *International Journal of Internet Protocol Technology*, *9*(1), 44. https://doi.org/10.1504/IJIPT.2015.074339

Baum, S. D. (2020). Social choice ethics in artificial intelligence. *AI & SOCIETY*, *35*(1), 165–176. https://doi.org/10.1007/s00146-017-0760-1

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Biddle, B. J. (1979). *Role Theory: Expectation, Identites, and Behaviors*. Elsevier. https://doi.org/10.1016/C2009-0-03121-3

Borbély, A., & Caputo, A. (2017). Approaching Negotiation at the Organizational Level. *Negotiation and Conflict Management Research*, *10*(4), 306–323. https://doi.org/10.1111/ncmr.12106

Brett, J. M. (2000). Culture and Negotiation. *International Journal of Psychology*, *35*(2), 97–104. https://doi.org/10.1080/002075900399385

Brett, J. M. (2002). Negotiating Globally: How to Negotiate Deals, Resolve Disputes, and Make Decisions across Cultural Boundaries. *International Marketing Review*, *19*(2), 206–208. https://doi.org/10.1108/imr.2002.19.2.206.2

Brett, J. M. (2017). Culture and negotiation strategy. *Journal of Business & Industrial Marketing*, *32*(4), 587–590. https://doi.org/10.1108/JBIM-11-2015-0230

Broekens, J., Jonker, C. M., & Meyer, J.-J. Ch. (2010). Affective negotiation support systems. *Journal of Ambient Intelligence and Smart Environments*, *2*(2), 121–144. https://doi.org/10.3233/AIS-2010-0065

Burr, C., Cristianini, N., & Ladyman, J. (2018). An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds and Machines*, *28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Butler, J. K. (1999a). Trust Expectations, Information Sharing, Climate of Trust, and Negotiation Effectiveness and Efficiency. *Group & Organization Management*, *24*(2), 217–238. https://doi.org/10.1177/1059601199242005

Butler, J. K. (1999b). Trust Expectations, Information Sharing, Climate of Trust, and Negotiation Effectiveness and Efficiency. *Group & Organization Management*, *24*(2), 217–238. https://doi.org/10.1177/1059601199242005

Cao, M., Hu, Q., Kiang, M. Y., & Hong, H. (2020). A Portfolio Strategy Design for Human-Computer Negotiations in e-Retail. *International Journal of Electronic Commerce*, *24*(3), 305–337. https://doi.org/10.1080/10864415.2020.1767428

Cao, M., Luo, X., Luo, X. (Robert), & Dai, X. (2015). Automated negotiation for e-commerce decision making: A goal deliberated agent architecture for multi-strategy selection. *Decision Support Systems*, *73*, 1–14. https://doi.org/10.1016/j.dss.2015.02.012

Cardot, H. (2011). *Recurrent Neural Networks for Temporal Data Processing*. https://api.semanticscholar.org/CorpusID:63146156

Castets-Renard, C. (2020). AI and the Law in the EU and the US. *Social Science Research Network*. https://consensus.app/papers/ai-and-the-law-in-the-eu-and-the-us-castets-renard/35f86074bd205d6386aa62a848157f35/

Chakraborty, S., Baarslag, T., & Kaisers, M. (2020). Automated peer-to-peer negotiation for energy contract settlements in residential cooperatives. *Applied Energy*, *259*, 114173. https://doi.org/10.1016/j.apenergy.2019.114173

Chasalow, K., & Levy, K. (2021). Representativeness in Statistics, Politics, and Machine Learning. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 77–89. https://doi.org/10.1145/3442188.3445872

Chen, S., & Su, R. (2022). An autonomous agent for negotiation with multiple communication channels using parametrized deep Q-network. *Mathematical Biosciences and Engineering*, *19*(8), 7933–7951. https://doi.org/10.3934/mbe.2022371

Chen, S., & Weiss, G. (2012). An efficient and adaptive approach to negotiation in complex environments. *Frontiers in Artificial Intelligence and Applications*, *242*, 228–233. https://doi.org/10.3233/978-1-61499-098-7-228

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005a). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In D. A. Moore (Ed.), *Conflicts of Interest: Challenges and Solutions in Business, Law, Medicine, and Public Policy* (pp. 74–95). Cambridge University Press.

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005b). Bounded Ethicality as a Psychological Barrier to Recognizing Conflicts of Interest. In *Conflicts of Interest* (pp. 74–95). Cambridge University Press. https://doi.org/10.1017/CBO9780511610332.006

Cohen, I. G., Amarasingham, R., Shah, A., Xie, B., & Lo, B. (2014). The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. *Health Affairs*, *33*(7), 1139–1147. https://doi.org/10.1377/hlthaff.2014.0048

Davidson, T. R., Veselovsky, V., Josifoski, M., Peyrard, M., Bosselut, A., Kosinski, M., & West, R. (2024). *Evaluating Language Model Agency through Negotiations*.

de Jong, K. A., Spears, W. M., & Gordon, D. F. (1993). Using Genetic Algorithms for Concept Learning. *Machine Learning*, *13*(2/3), 161–188. https://doi.org/10.1023/A:1022617912649

de Melo, C. M., Marsella, S., & Gratch, J. (2016). “Do As I Say, Not As I Do”: Challenges in Delegating Decisions to Automated Agents. *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems*, 949–956.

de Melo, C. M., Marsella, S., & Gratch, J. (2018). Social decisions and fairness change when people’s interests are represented by autonomous agents. *Autonomous Agents and Multi-Agent Systems*, *32*(1), 163–187. https://doi.org/10.1007/s10458-017-9376-6

Deng, B. (2015). Machine ethics: The robot’s dilemma. *Nature*, *523*(7558), 24–26. https://doi.org/10.1038/523024a

Dimopoulos, Y., & Moraitis, P. (2014). Advances in Argumentation-Based Negotiation. In Y. Dimopoulos & P. Moraitis (Eds.), *Negotiation and Argumentation in Multi-Agent Systems* (pp. 82–125). BENTHAM SCIENCE PUBLISHERS. https://doi.org/10.2174/978160805824211401006

DiPietro, R., & Hager, G. D. (2020). Deep learning: RNNs and LSTM. In *Handbook of Medical Image Computing and Computer Assisted Intervention* (pp. 503–519). Elsevier. https://doi.org/10.1016/B978-0-12-816176-0.00026-0

Doğru, A., Keskin, M. O., & Aydoğan, R. (2025). Taking into Account Opponent’s Arguments in Human-Agent Negotiations. *ACM Transactions on Interactive Intelligent Systems*, *15*(1), 1–35. https://doi.org/10.1145/3691643

Dommen, E. C., Walton, R. E., & McKersie, R. B. (1966). A Behavioral Theory of Labor Negotiations. *The Economic Journal*, *76*(302), 406. https://doi.org/10.2307/2229740

Duin, A. H., & Pedersen, I. (2021). Working Alongside Non-Human Agents. *2021 IEEE International Professional Communication Conference (ProComm)*, 1–5. https://doi.org/10.1109/ProComm52174.2021.00005

Elms, D. (2006). How Bargaining Alters Outcomes: Bilateral Trade Negotiations and Bargaining Strategies. *International Negotiation*, *11*(3), 399–429. https://doi.org/10.1163/157180606779155237

Falcão Filho, H. A. (2024). Making sense of negotiation and AI: The blossoming of a new collaboration. *International Journal of Commerce and Contracting*, *8*(1–2), 44–64. https://doi.org/10.1177/20555636241269270

Fernandez, A., Garcia, S., Luengo, J., Bernado-Mansilla, E., & Herrera, F. (2010). Genetics-Based Machine Learning for Rule Induction: State of the Art, Taxonomy, and Comparative Study. *IEEE Transactions on Evolutionary Computation*, *14*(6), 913–941. https://doi.org/10.1109/TEVC.2009.2039140

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. *Minds and Machines*, *28*(4), 689–707. https://doi.org/10.1007/s11023-018-9482-5

Franks, E., Lee, B., & Xu, H. (2024). Report: China's New AI Regulations. *Global Privacy Law Review*, *5*(Issue 1), 43–49. https://doi.org/10.54648/GPLR2024007

Freeman, R. E. E., & McVea, J. (2001). A Stakeholder Approach to Strategic Management. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.263511

Fuchs, A., Passarella, A., & Conti, M. (2023). Modeling, Replicating, and Predicting Human Behavior: A Survey. *ACM Transactions on Autonomous and Adaptive Systems*, *18*(2), 1–47. https://doi.org/10.1145/3580492

Georgiadis, G., Ravid, D., & Szentes, B. (2024). Flexible Moral Hazard Problems. *Econometrica*, *92*(2), 387–409. https://doi.org/10.3982/ECTA21383

Gill, T. (2020). Blame It on the Self-Driving Car: How Autonomous Vehicles Can Alter Consumer Morality. *Journal of Consumer Research*, *47*(2), 272–291. https://doi.org/10.1093/jcr/ucaa018

Goldberg, Y. (2017). Neural Network Methods for Natural Language Processing. *Synthesis Lectures on Human Language Technologies*, *10*(1), 16–20. https://doi.org/10.2200/S00762ED1V01Y201703HLT037

Grossberg, S. (2013). Recurrent neural networks. *Scholarpedia*, *8*(2), 1888. https://doi.org/10.4249/scholarpedia.1888

Gunia, B. C., Brett, J. M., Nandkeolyar, A. K., & Kamdar, D. (2011). Paying a price: Culture, trust, and negotiation consequences. *Journal of Applied Psychology*, *96*(4), 774–789. https://doi.org/10.1037/a0021986

Habumuremyi, J. B., & K Tarus, T. (2021). Effect of Stakeholders' Participation on Sustainability of Community Projects in Ruhango District, Rwanda. *International Journal of Research and Innovation in Social Science*, *05*(09), 429–433. https://doi.org/10.47772/IJRISS.2021.5926

Haim, G., Gal, Y. (Kobi), An, B., & Kraus, S. (2017). Human–computer negotiation in a three player market setting. *Artificial Intelligence*, *246*, 34–52. https://doi.org/10.1016/j.artint.2017.01.003

Hannecke, M. (2024, November 20). *CRISP-DM in AI/ML An overview | Bluetuple.ai*. Medium. https://medium.com/bluetuple-ai/enhancing-ai-fairness-crisp-dm-in-ai-ml-and-the-bias-blindspot-ad9cfdbe5e53

High-Level Expert Group on AI. (2019). *Ethics Guidelines for Trustworthy AI - High-Level Expert Group on Artificial Intelligence*. https://ec.europa.eu/digital-

Hunter, A. (2015). *Modelling the persuadee in asymmetric argumentation dialogues for persuasion*.

Jennings, N. R., Faratin, P., Lomuscio, A. R., Parsons, S., Wooldridge, M. J., & Sierra, C. (2001). Automated Negotiation: Prospects, Methods and Challenges. *Group Decision and Negotiation*, *10*(2), 199–215. https://doi.org/10.1023/A:1008746126376

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2

Jones, T. M. (1991). Ethical Decision Making by Individuals in Organizations: An Issue-Contingent Model. *The Academy of Management Review*, *16*(2), 366. https://doi.org/10.2307/258867

Jonker, C. M., Hindriks, K. V, Wiggers, P., & Broekens, J. (2012). *Negotiating Agents*.

Kersten, G. E., & Lai, H. (2007). Negotiation Support and E-negotiation Systems: An Overview. *Group Decision and Negotiation*, *16*(6), 553–586. https://doi.org/10.1007/s10726-007-9095-5

Kloe, R., Zylowski, T., & Zirpins, C. (2020). Dynamic Re-Configuration of Conversationally Initiated Automated Negotiations. *2020 IEEE 24th International Enterprise Distributed Object Computing Workshop (EDOCW)*, 95–98. https://doi.org/10.1109/EDOCW49879.2020.00026

Kong, D. T., Dirks, K. T., & Ferrin, D. L. (2014). Interpersonal Trust within Negotiations: Meta-Analytic Evidence, Critical Contingencies, and Directions for Future Research. *Academy of Management Journal*, *57*(5), 1235–1255. https://doi.org/10.5465/amj.2012.0461

Kong, D. T., & Yao, J. (2019). Advancing the Scientific Understanding of Trust and Culture in Negotiations. *Negotiation and Conflict Management Research*, *12*(2), 117–130. https://doi.org/10.1111/ncmr.12147

Kröhling, D. E., Chiotti, O. J. A., & Martínez, E. C. (2024). Context-Aware Cognitive Agents using Knowledge Graphs for Automated Negotiation. *2024 L Latin American Computer Conference (CLEI)*, 1–10. https://doi.org/10.1109/CLEI64178.2024.10700165

Lacher, S. R. (1981). LEGAL CONSIDERATIONS AND NEGOTIATION OF JOINT EXPLORATION AGREEMENTS. *The APPEA Journal*, *21*(1), 41. https://doi.org/10.1071/AJ80005

Lee, C. C., Rahiman, M. H. F., Rahim, R. A., & Saad, F. S. A. (2021). A Deep Feedforward Neural Network Model for Image Prediction. *Journal of Physics: Conference Series*, *1878*(1), 012062. https://doi.org/10.1088/1742-6596/1878/1/012062

Li, H., Ni, T., Agrawal, S., Jia, F., Raja, S., Gui, Y., Hughes, D., Lewis, M., & Sycara, K. (2021). Individualized Mutual Adaptation in Human-Agent Teams. *IEEE Transactions on Human-Machine Systems*, *51*(6), 706–714. https://doi.org/10.1109/THMS.2021.3107675

Liang, T., Farrell, M. H., & Misra, S. (2020). *DEEP NEURAL NETWORKS FOR ESTIMATION AND INFERENCE*. https://api.semanticscholar.org/CorpusID:267861888

Lim, S. G.-S., & Murnighan, J. K. (1994). Phases, Deadlines, and the Bargaining Process. *Organizational Behavior and Human Decision Processes*, *58*(2), 153–171. https://doi.org/10.1006/obhd.1994.1032

Lopes, F., Wooldridge, M., & Novais, A. Q. (2008). Negotiation among autonomous computational agents: principles, analysis and challenges. *Artificial Intelligence Review*, *29*(1), 1–44. https://doi.org/10.1007/s10462-009-9107-8

Lügger, K., Geiger, I., Neun, H., & Backhaus, K. (2015). When East meets West at the bargaining table: adaptation, behavior and outcomes in intra- and intercultural German–Chinese business negotiations. *Journal of Business Economics*, *85*(1), 15–43. https://doi.org/10.1007/s11573-013-0703-3

Macdonell, S. G., Gibb, D. J., & Dowdeswell, B. (2021). *DIAGNOSTIC BELIEF-DESIRE-INTENTION AGENTS FOR DISTRIBUTED IEC 61499 FAULT DIAGNOSIS*.

Malle, B. F. (2016). Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*, *18*(4), 243–256. https://doi.org/10.1007/s10676-015-9367-8

Manchev, A. (2024). WORLD'S FIRST LAW FOR ARTIFICIAL INTELLIGENCE. LEGAL, ETHICAL AND ECONOMIC ASPECTS. *Education and Technologies Journal*, *15*(1), 225–228. https://doi.org/10.26883/2010.241.5985

Martin, K. (2019). Ethical Implications and Accountability of Algorithms. *Journal of Business Ethics*, *160*(4), 835–850. https://doi.org/10.1007/s10551-018-3921-3

Mell, J., Beissinger, M., & Gratch, J. (2021). An expert-model and machine learning hybrid approach to predicting human-agent negotiation outcomes in varied data. *Journal on Multimodal User Interfaces*, *15*(2), 215–227. https://doi.org/10.1007/s12193-021-00368-w

Mell, J., Lucas, G., Mozgai, S., & Gratch, J. (2020). The Effects of Experience on Deception in Human-Agent Negotiation. *Journal of Artificial Intelligence Research*, *68*, 633–660. https://doi.org/10.1613/jair.1.11924

Memon, M. A., Scoccia, G. L., Autili, M., & Inverardi, P. (2024). An Architecture for Ethics-Based Negotiation in the Decision-Making of Intelligent Autonomous Systems. *2024 IEEE 21st International Conference on Software Architecture Companion (ICSA-C)*, 60–64. https://doi.org/10.1109/ICSA-C63560.2024.00016

Mertes, M., Kunz, D., & Hüffmeier, J. (2023). A Deep Dive Into Distributive Concession Making and the Likelihood of Impasses in Negotiations. *Collabra: Psychology*, *9*(1). https://doi.org/10.1525/collabra.88929

Miller, G. J. (2019). Quantitative Comparison of Big Data Analytics and Business Intelligence Project Success Factors. In E. Ziemba (Ed.), *Information Technology for Management: Emerging Research and Applications* (pp. 53–72). Springer International Publishing.

Miller, G. J. (2022). Stakeholder roles in artificial intelligence projects. *Project Leadership and Society*, *3*, 1–15. https://doi.org/10.1016/j.plas.2022.100068

Mirzayi, S., Taghiyareh, F., & Nassiri-Mofakham, F. (2022). An opponent-adaptive strategy to increase utility and fairness in agents' negotiation. *Applied Intelligence*, *52*(4), 3587–3603. https://doi.org/10.1007/s10489-021-02638-2

Mohammad, Y., & Nakadai, S. (2018). *FastVOI: Efficient Utility Elicitation During Negotiations* (pp. 560–567). https://doi.org/10.1007/978-3-030-03098-8_42

Moser, C., den Hond, F., & Lindebaum, D. (2022). Morality in the Age of Artificially Intelligent Algorithms. *Academy of Management Learning & Education*, *21*(1), 139–155. https://doi.org/10.5465/amle.2020.0287

OECD. (2019). *Artificial Intelligence in Society*. OECD Publishing. https://doi.org/10.1787/eedfee77-en

Okamura, K., & Yamada, S. (2020). Adaptive trust calibration for human-AI collaboration. *PLOS ONE*, *15*(2), e0229132. https://doi.org/10.1371/journal.pone.0229132

Oreski, D., & Androcec, D. (2020). Genetic algorithm and artificial neural network for network forensic analytics. *2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1200–1205. https://doi.org/10.23919/MIPRO48935.2020.9245140

Oshrat, Y., Lin, R., & Kraus, S. (2009). Facing the challenge of human-agent negotiations via effective general opponent modeling. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS* (Vol. 2). https://doi.org/10.1145/1558013.1558064

Papaioannou, I. V., Roussaki, I. G., & Anagnostou, M. E. (2008). Neural networks against genetic algorithms for negotiating agent behaviour prediction. *Web Intelligence and Agent Systems: An International Journal*, *6*(2), 217–233. https://doi.org/10.3233/WIA-2008-0138

Park, J., Rahman, H. A., Suh, J., & Hussin, H. (2019a). A Study of Integrative Bargaining Model with Argumentation-Based Negotiation. *Sustainability*, *11*(23), 6832. https://doi.org/10.3390/su11236832

Park, J., Rahman, H. A., Suh, J., & Hussin, H. (2019b). A Study of Integrative Bargaining Model with Argumentation-Based Negotiation. *Sustainability*, *11*(23), 6832. https://doi.org/10.3390/su11236832

Peterson, R. M., & Lucas, G. H. (2001). Expanding the Antecedent Component of the Traditional Business Negotiation Model: Pre-Negotiation Literature Review and Planning-Preparation Propositions. *Journal of Marketing Theory and Practice*, *9*(4), 37–49. https://doi.org/10.1080/10696679.2001.11501902

Phogat, K. S., & Tsumura, K. (2024). Dynamic Programming Principle for Automatic Negotiations. *IEEE Transactions on Automatic Control*, *69*(2), 1281–1287. https://doi.org/10.1109/TAC.2023.3285859

Picard, R. W., & Picard, R. (1997). *Affective computing* (Vol. 252). MIT press. https://doi.org/10.1037/e526112012-054

Pienaar, W. D., & Spoelstra, H. I. (1996). *Negotiation: Theories, Strategies, and Skills*. Juta and Company Ltd.

Pinkley, R. L., & Northcraft, G. B. (1994). Conflict Frames of Reference: Implications for Dispute Processes and Outcomes. *Academy of Management Journal*, *37*(1), 193–205. https://doi.org/10.5465/256777

PMI, P. M. I. (2021). *A Guide to the Project Management Body of Knowledge (PMBOK Guide) Seventh Edition* (7th ed.). Project Management Institute.

Pruitt, D. G., & Lewis, S. A. (1975). Development of integrative solutions in bilateral negotiation. *Journal of Personality and Social Psychology*, *31*(4), 621–633. https://doi.org/10.1037/0022-3514.31.4.621

Rajavel, R., & Thangarathanam, M. (2016). Adaptive Probabilistic Behavioural Learning System for the effective behavioural decision in cloud trading negotiation market. *Future Generation Computer Systems*, *58*, 29–41. https://doi.org/10.1016/j.future.2015.12.007

Ren, W., Wu, J., Zhang, X., Lai, R., & Chen, L. (2018). A Stochastic Model of Cascading Failure Dynamics in Communication Networks. *IEEE Transactions on Circuits and Systems II: Express Briefs*, *65*(5), 632–636. https://doi.org/10.1109/TCSII.2018.2822049

Rinehart, L. M., & Page, T. J. (1992). The Development and Test of a Model of Transaction Negotiation. *Journal of Marketing*, *56*(4), 18–32. https://doi.org/10.1177/002224299205600403

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not So Different After All: A Cross-Discipline View Of Trust. *Academy of Management Review*, *23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Ruslin, R., Mashuri, S., Sarib, M., Alhabsyi, F., & Syam, H. (2022). *Semi-structured Interview: A Methodological Reflection on the Development of a Qualitative Research Instrument in Educational Studies Ruslin*. *Vol. 12*, 22–29. https://doi.org/10.9790/7388-1201052229

Ryan, M., & Stahl, B. C. (2021). Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications. *Journal of Information, Communication and Ethics in Society*, *19*(1), 61–86. https://doi.org/10.1108/JICES-12-2019-0138

Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). Re-imagining Algorithmic Fairness in India and Beyond. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 315–328. https://doi.org/10.1145/3442188.3445896

Sharma, P., Malik, N., Akhtar, N., & Rohilla, H. (2013). *FEEDFORWARD NEURAL NETWORK: A Review*. https://api.semanticscholar.org/CorpusID:62442434

Simon, J. P. (2019). Artificial intelligence: scope, players, markets and geography. *Digital Policy, Regulation and Governance*, *21*(3), 208–237. https://doi.org/10.1108/DPRG-08-2018-0039

Steinel, W., & Harinck, F. (2020). Negotiation and Bargaining. In *Oxford Research Encyclopedia of Psychology*. Oxford University Press. https://doi.org/10.1093/acrefore/9780190236557.013.253

Stevens, C. A., Daamen, J., Gaudrain, E., Renkema, T., Top, J. D., Cnossen, F., & Taatgen, N. A. (2018). Using Cognitive Agents to Train Negotiation Skills. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00154

Sticher, V. (2021). Negotiating Peace with Your Enemy: The Problem of Costly Concessions. *Journal of Global Security Studies*, *6*(4). https://doi.org/10.1093/jogss/ogaa054

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions? *Trends in Cognitive Sciences*, *19*(1), 27–34. https://doi.org/10.1016/j.tics.2014.11.005

Sun, Q., Zhang, X., Jin, R., Zhang, X., & Ma, Y. (2024). Multi-strategy synthetized equilibrium optimizer and application. *PeerJ Computer Science*, *10*, e1760. https://doi.org/10.7717/peerj-cs.1760

Tovar Cardozo, G. (2023). Approach to global regulations around AI. *LatIA*, *1*, 7. https://doi.org/10.62486/latia20237

Vesa, M., & Tienari, J. (2022). Artificial intelligence and rationalized unaccountability: Ideology of the elites? *Organization*, *29*(6), 1133–1145. https://doi.org/10.1177/1350508420963872

Vij, S. R., More, A., Mukhopadhyay, D., & Agrawal, A. J. (2015). An E-Negotiation Agent Using Rule Based and Case Based Approaches: A Comparative Study with Bilateral E-Negotiation with Prediction. *Journal of Software Engineering and Applications*, *08*(10), 521–530. https://doi.org/10.4236/jsea.2015.810049

Vos, J. F. J., & Achterkamp, M. C. (2006). Stakeholder identification in innovation projects. *European Journal of Innovation Management*, *9*(2), 161–178. https://doi.org/10.1108/14601060610663550

Weingart, L. R., Brett, J. M., Olekalns, M., & Smith, P. L. (2007). Conflicting social motives in negotiating groups. *Journal of Personality and Social Psychology*, *93*(6), 994–1010. https://doi.org/10.1037/0022-3514.93.6.994

Weingart, L. R., Thompson, L. L., Bazerman, M. H., & Carroll, J. S. (1990). TACTICAL BEHAVIOR AND NEGOTIATION OUTCOMES. *International Journal of Conflict Management*, *1*(1), 7–31. https://doi.org/10.1108/eb022670

Wieringa, M. (2020). What to account for when accounting for algorithms. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 1–18. https://doi.org/10.1145/3351095.3372833

Yalçın, O. G. (2021). Feedforward Neural Networks. In *Applied Neural Networks with TensorFlow 2* (pp. 121–143). Apress. https://doi.org/10.1007/978-1-4842-6513-0_6

Yang, L., Allen, G., Zhang, Z., & Zhao, Y. (2024). Achieving On-Site Trustworthy AI Implementation in the Construction Industry: A Framework Across the AI Lifecycle. *Buildings*, *15*(1), 21. https://doi.org/10.3390/buildings15010021

Yang, Y., Singhal, S., & Xu, Y. (2009). Offer with Choices and Accept with Delay: A Win-Win Strategy Model for Agent Based Automated Negotiation. In *ICIS 2009 Proceedings - Thirtieth International Conference on Information Systems*.

Yao, J., Zhang, Z., & Brett, J. M. (2017). Understanding trust development in negotiations: An interdependent approach. *Journal of Organizational Behavior*, *38*(5), 712–729. https://doi.org/10.1002/job.2160

Yu, B., Vahidov, R., & Kersten, G. E. (2021). Acceptance of technological agency: Beyond the perception of utilitarian value. *Information & Management*, *58*(7), 103503. https://doi.org/10.1016/j.im.2021.103503

Zarghami, S. A., & Dumrak, J. (2021). Reimaging stakeholder analysis in project management: network theory and fuzzy logic applications. *Engineering, Construction and Architectural Management*, *28*(9), 2426–2447. https://doi.org/10.1108/ECAM-06-2020-0391

Zhang, R., Shamma, J., & Li, N. (2024). *Equilibrium Selection for Multi-agent Reinforcement Learning: A Unified Framework*.

Zwikael, O., & Meredith, J. R. (2018). Who's who in the project zoo? The ten core project roles. *International Journal of Operations & Production Management*, *38*(2), 474–492. https://doi.org/10.1108/IJOPM-05-2017-0274

(page left purposely blank)

Appendices

Appendix A – Stakeholder Roles defined by Miller (2022)

Stakeholder (Project role)	Roles by Miller (2022)
Platform Owner	"The technology platform is the technical infrastructure and architecture upon which the AI system resides. It can be owned and hosted by the platform owner, which means that it could be owned and hosted in-house or in the cloud by the operating organization, or it could be hosted on online platforms managed by third-party providers such as cloud service providers, social media sites, or search engines."
Functional Departments	-
Data Subjects Decision Subjects Workers	"… include the people who may have an agreement with the development organization or the operating organization, such as terms of services or service or employment contracts."
End Users	"… include individuals engaged in system usage through employment or service contracts with the operating organization or as consumers. They may have direct system access or receive the system outputs through an intermediary user."
Advocates	"… are third parties who can represent different parties with the development organization, the operating organization, or individuals and groups in the AI community. This group is not harmed by the system usage or development and may have coercive power over the project."
Independent Evaluators	"… include third parties such as investigative reporters in the media, journalists, internal or external auditors, safety certifiers, institutional review boards, insurers, judges, lawyers, or third-party reviewers such as the general public. Depending on the engagement model, these parties may have coercive or reward power with the development organization or the operating organization."
Courts and Regulators	"… enforce compliance with regulations, laws, acts, and ordinances."
Model Maintainers	"Model values or choices can become obsolete. They must be reviewed and assessed for fitness and refreshed over time. This action is performed by model maintainers."
Data Controller (data custodian)	"… is responsible for data governance and its use in system development."
Legal and Ethical Functions	"… include supervisors, boards, or internal support roles (e.g., whistle-blower, ombudsman) that provide needful guidance, policies, and support for ethical practices and issues."
Executive Management	"… includes senior and top management support for the project manager."
Project Funder	"… is the executive management who commits the resources (labour or money, for example) to the project. Multiple funders may become involved between the development and usage stages, and funders do not play an active role in the project."
Project Owner	"… provides strategic direction to the project, defining

	the project scope, arranging the performing entity, and chairing the steering committee. Thus, the project owner issues the statement of work (SOW), the scope definition document, or the problem statement that defines the aims and rationale for the system."
Project Manager	"… is accountable for the project's outputs per an agreed SOW provided by the project owner. Moreover, the project manager ensures adherence to ethical, privacy, and security norms and expectations and the engagement of the relevant stakeholders."
Business Users	"… are responsible for specific content development, editing, or coaching the team or users."
Data Analysts	"… include domain experts, industry experts, analysts, and consultants who understand processes and government policy; they have expertise in the context needed to interpret raw data, ask relevant questions, generate relevant hypotheses and interpret the results."
Data Scientists/AI Practitioners	"… is an interdisciplinary role that aims to transform data into knowledge by finding patterns and trends in the data. However, the role of an AI practitioner (e.g., computer vision engineer) may be relevant, depending on the technology and context of the project."
Data Engineers	"… is responsible for data preparation and quality, including extracting, cleaning, enriching, and transforming the data. This role may be directly or indirectly responsible for annotating or labelling data within a dataset. Data engineers may include curators, data collectors, data custodians, data experts, and data stewards."
Software Developers	"… includes computer scientists who use software tools and technologies, programming language, and technical infrastructure to implement and optimize the processing of large data sets or developing user interfaces. Software development may include report visualization developers, designers, and other developers."
Software Architects	"… is responsible for technical infrastructure and may include architects, the IT team, the process manager, and security specialists."

Appendix B – Answers given for the multiple interview questions

Appendix B.1 - Answers provided for the Interview Question 1

Interview Question 1: Who are the stakeholders during the data collection and preparation process?	
1	Ok. Por exemplo, o software arquitect, se bem que pode ser sempre feito ajustos, mas à partida esse estudo da arquitetura já foi feito antes dos developers estarem a trabalhar, I guess. Ou então eu estou a dar uma resposta muito com base no que é o meu projeto atual. Ou seja, já está uma arquitetura definida. E nós ao longo do tempo vamos trabalhando. Pode haver ligeiros e ajustos. Mas tipo, arquitetura, tipo, o bolo em si já está mais ou menos definido. Mas eu acho que os developers acabam por ser sempre as pessoas que acabam por estar sempre com as mãos na massa.
2	Ok, então é mais os data scientists e... Os Data Engineers, porque os Data Analysts acredito que entrem um bocadinho mais tarde. (Data Subjects) Ou seja, são as pessoas pelas quais eu tenho os dados, é isso? Se sim, também fazem parte. Tecnicamente são um bocadinho stakeholders, sim, sendo que a informação é deles. Eu acho que em qualquer fase, é sempre o dono do projeto lá. Acho que vai sempre ser stakeholder para mim em todas as fases. Porque é o mais interessado no fundo disto tudo.
3	Tens o Manager, depois tens o Product Owner, depois ainda tens em cima... Já deve ser mais a nível de empresa já, não é. Então, eu diria que tens que ter uma equipa de levantamento inicial, não é? Essa equipa inicial tem que ter pessoas de várias valências, tens de ter algumas mais funcionais, outras mais técnicas. Sim, mas dos que eu estou a ver tens aí, tens os data analistas, tens a parte dos arquitetos, tens os developers que se complementam, tens a parte dos engenheiros, os data engineering que vai muito também por causa da inteligência artificial, eu diria que anda por aí.
4	Começar a dizer, no fundo qualquer pessoa envolvida no projeto do lado do cliente e do lado da empresa subcontratada eu diria que estão envolvidos porque vão lidar com os dados e diria que é só. Exato.
5	Data engineers, se calhar vão começar inicialmente a pegar nos dados mas estou a pensar primeiro na fase de escolher que dados em que queremos pegar aí se calhar pegam mais nos data scientists, business users mais no sentido de decidir o end product. (...) data scientists, data engineers, business users em cima, depois project managers que estão nas decisões dos baixos, data controller, project owner diria que estão mais dentro das decisões de que dados é que querem e que dados é que precisam. Precisam da aprovação do Project Funder.
6	há uma fase antes, e depois então o Source Data, é que são mais o Project Team e o Project Manager e o Data Controller e o Project Owner, que vão ver quais são tipicamente os dados que existem, e o Legal and Ethical Functions pode entrar aqui, porque pode haver acordos de Confidencialidade de dados e Confidencialidade de anonimização que têm que ser feitos para tudo avançar. Os DPO's, portanto, os encarregados de proteção de dados das diferentes organizações.
7	The data collection should be the project manager, the product owner, and the user, the client, in my opinion. They should prepare the data, anything that should be used to relevance for the developers, for them to work. Basically, the project owner manager is in between the client and the developer, the team of developers. Okay, in my opinion, from what I saw and from what I've learned, I thought the client should not be directly in contact with the team of developers, okay, because any demand, any need should first be go through the product owner, so that he can understand, organize, qualify the needs, then give them the task to the developer. Because if that, because the product manager should be the one that have a whole view of the project, and if there's a need that seems to not be like giving any value, adding value to the project, he's the one who should just say no, it's not possible to do that, or I have to talk to the developer first and then people, that's the client's.
8	(..) colega meu sim, sim. Ele sim utiliza muito essa parte. Ele é cientista de dados e portanto é quem trabalha com esses dados.
9	Normalmente, há sempre um gestor de projeto que há de ter falado com o cliente primeiro, não é? Que há de ter um objectivo. Depois, normalmente, esse gestor de projeto há de entrar em contacto com alguém da parte mais técnica. A propor o que é que vai ser executado. Às vezes, é essa parte técnica que decide as fontes dos dados. Ou temos o cliente que diz, eu já tenho um conjunto de dados que quero utilizar. Isto depois também depende muito do que é que é o objectivo do desenvolvimento. É para o que podemos pressupor. Imagina que temos um agente que é para trabalhar com dados internos na empresa. Portanto, vão ser documentos que o cliente vai entregar, portanto não vão ser documentos de uma entidade externa. Portanto vais treinar o modelo com os documentos que foram entregados e é essa a informação que só entra ali dentro. Há outro tipo de softwares, outro tipo the LEMs, há mais que um, que imagina que é um software de faturação. Aí já é a própria empresa que vai buscar dados externos, com o máximo de informação externa, porque o cliente tem muito pouco. Aí seriam as principais fontes e está muito dependente daquilo que é o objetivo, neste caso, do algoritmo. No nosso caso, as nossas fontes, essencialmente, são sensores. Nós temos um conjunto de sensores que aplicamos. Por exemplo. A nível de uma cidade, as estações meteorológicas, sensores de CO2, até às vezes pode ser sensores de humidade do solo que está nos jardins. A gente faz essa agregação mais deste tipo de sensores e alimentamos o modelo. Podemos fazer várias coisas com ele, mas essencialmente, no nosso caso, seriam mais sensores a fonte de dados.
10	Idealmente sim e em certas situações até faz sentido que o primeiro contato com o cliente seja logo feito com o data analyst ao lado. Na minha ótica até é recomendável que esse tipo de pessoas faça parte de todas as fases do projeto obviamente que eles não vão fazer a orçamentação, nem é isso que se lhes pede mas participam nas reuniões com o cliente para darem o seu parecer logo na hora da área onde são especialistas. Pronto eu diria que aí será um conjunto de pessoas, portanto, tanto da parte dos gestores de topo que têm que dar o seu aval mas tem que haver também um contacto com os data scientists e o pessoal, (...) mas depois quem dará o certo final acho que passará sempre por ser a parte da engenharia de dados.
11	Sim, pois é que eu não sei qual é o nível de detalhes, se isto tem algum CRISP-DM por trás. Não sei se conhece a arquitetura de CRISPM, do data mining, ou seja, nós temos uma ideia de negócio, temos os requisitos bastante bem definidos, qual é a melhor solução? Temos que ir ver, temos que ir fazer um levantamento daquilo que já existe para que possamos escolher a tecnologia, escolher o algoritmo e aí sim. Percorrer aquilo que diz aqui, source data, data collection, training. Portanto ainda tem que ser feito uma espécie de identificação de melhores tecnologias melhores frameworks, melhores bibliotecas, melhores algoritmos, de acordo com o nosso caso de estudo. E dar a nossa solução final.

Appendix B.2 - Answers provided for the Interview Question 2

	Interview Question 2: What are the responsibilities of the stakeholders (e.g., data engineers, analysts) in ensuring data quality and privacy?
1	Às vezes existe uma entidade só para isso. Por exemplo, a Deloitte sabendo que vai haver auditorias, contrata ela um serviço de auditoria para garantir que está tudo direito. Pode ser alguma coisa interna também. Ou seja, faz uma pré-auditoria antes de uma auditoria real.
2	O owner do projeto e da empresa, no geral, tem responsabilidade porque é o chefe máximo, não é? E é ele que está a encomendar o projeto, portanto é ele sempre que vai ter a maior responsabilidade. A seguir, o project manager. Primeiro a extração dos dados tem que ser feita consoante as regras que existem. Existem até normas da União Europeia, portanto não há muito por onde fugir. E depois, a própria manipulação dos dados também tem regras. E, acima de tudo, eu não disse o software arquiteto mas talvez o software, não sei em que processo é ele está, mas, ou seja, quem está encarregue de tratar do armazenamento dos dados, que eu acredito que não seja nenhum destes não seja o data scientist, nem o data engineer, seria mais um engenheiro de sistemas, talvez, os engenheiros de sistemas também têm que garantir que está seguro dos dados. Mas ele não está aqui, penso.
3	Tendo em conta que o trabalho deles é sensível, sim, mas eles assinam um contrato. Por outro, o nosso trabalho também, quando se partilha informação, é anonimizá-la ao máximo. Quando se partilha informação que tenha dados sensíveis, nós temos a obrigatoriedade enquanto trabalhadores diretos. Eles como clientes também têm a sua responsabilidade no contrato, mas quando há partilha de informação pronto, mete-se aquele planos desfocados e outras coisas para não poderem ter acesso às informações. Dá-se users test e ambientes de estudo em qualidade nunca se dá acesso a produção nem nada assim do género pronto, essas medidas têm que ser garantidas sempre. A base é por aí, depois se tiverem que ter acesso mesmo a alguma coisa, pronto, aí já tem que ser pronto, com esse cuidado.
4	Sim é assim, dependendo do tipo de dados que são, por exemplo há projetos em que os dados que são utilizados para desenvolvimento são dados de AMI, ou seja, não são dados reais de ambientes de produção. Porque desta forma se calhar os developers mais vulneráveis têm acesso a este tipo de informação. E porque o cliente não quer passar informação entre sistemas, entre ambientes, por exemplo, e depois há certas seguranças que depois são mantidas mesmo ao longo do desenvolvimento da plataforma, nomeadamente o lashing de passwords, a encriptação de dados, por exemplo para não ser possível visualizar determinados identificadores e passwords ao nível do inspector, por exemplo se for uma aplicação online, se é possível, para não estar visível no client-side. E se calhar também uma grande parte da responsabilidade do cliente para também fazer esse tipo de proteção porque a partir do momento que mesmo o desenvolvimento vai acontecendo com dados especiais, os developers, em parte alguns, terão sempre acesso aos ambientes que eram. Portanto, isso depois precisa de ser protegido, os tipos de acessos e tudo mais, e a forma como essa informação é retirada.
5	Eu acho que tem engineers e tem mais responsabilidades se calhar em garantir por exemplo questões de privacidade dos dados que não têm PIIs, Personal Identifiable Information. Sim, dessa primeira fila, eu acho que tem mais responsabilidade, diria que era isso.
6	E mais, e isto é uma questão fundamental de garantir que os dados são usados em cumprimento de todas as disposições de dados do Regulamento de Proteção de Dados. (...) por exemplo, nós no projeto de fraude trabalhámos com dados sensíveis. E os acordos de confidencialidade que tiveram de ser feitos levaram cerca de 10 meses.
7	Quality and privacy, the developers should ensure that the data are secured are encrypted and that are not you know modified or tempered in in the wrong way. Data shouldn't be modified but as intended because you have for example you have to change your formats or something like that that should not be modified just in just like in a way that are not useful or wrongly displayed for both clients but at the task of the developers. For the product managers they should ensure that to give the team of developers the right data that are needed for them to be able to work and to have access to this data. The right access and basically to understand the kind of overview of how the structure of the data works like what data come from where and involved with what. I think the project manager should be to know that if the if the one developer the developer asks them how where does this data come from? Where this number comes from? The project manager should know or should know where to find that answer.
8	
9	Neste caso, a qualidade não é supervisionada. Não tens ali uma pessoa ativamente a verificar a qualidade. Poderá haver uma anomalia, que a gente chega ali e vê que um sensor se está a comportar mal e detetamos essa anomalia e é corrigida, porque muitas das vezes esses dados já estão dentro do sistema. Normalmente, o que a gente está à espera quando faz um sistema de inteligência artificial é que ele também consiga lidar com esse tipo de anomalias e tens um pré-processamento dos dados, vá lá um filtro prévio, antes dos dados entrarem na base de dados, que evita que isso aconteça. Há várias técnicas para fazer isso. Uma delas é fazer as medianas dos valores para manter aqueles picos. Normalmente o que se faz é, tens o RGPD, as normas seguem o RGPD, quando envolve pessoas, claramente essas pessoas têm que ser informadas, quais são os dados que vão ser recolhidos e a gente faz o melhor possível para não haver mais nada entrar no sistema sem ser aquilo que as pessoas permitiram. Normalmente estamos a falar em meta dados, muito raramente, que se consegue identificar uma pessoa. E normalmente quando estás a fazer um algoritmo de inteligência artificial é isso que tu queres, é uma abstração, não queres um singular, queres ter a ideia geral do comportamento de várias pessoas. O interesse de uma coisa muito pessoal não tem interesse.
10	neste caso chamado da parte da analítica de dados por si vão ter que saber que dados é que vão tratar e eles é que são os técnicos da área, e eles é que deverão definir se faz sentido ou não tratar esses dados.
11	

Appendix B.3 - Answers provided for the Interview Question 3

	Interview Question 3: Who do you consider to be the stakeholders during the planning and design process?
1	Development, eu vou sempre dizer os developers, se calhar ali em conjunto com não sei bem business users, data analysts, data scientists AI. Acaba por haver sempre uma cooperação entre todos.
2	Então aqui continuam a ser as pessoas do projeto, aqui para além do Data Scientist e do Data Engineer, acredito que também já entra o Data Analyst. O developer acho que não, acho que ainda não, para esta fase ainda não. Talvez o Software Architect.
3	Ok, os mesmos têm que estar, os que forem envolvidos no levantamento têm que estar. Sim, e aí tens envolvido algum UX designer ou alguma coisa assim.
4	Em parte, se calhar não toda a equipa do lado do cliente, mas tem de haver alguém do lado do cliente sempre, uma pessoa ou um grupo de pessoas que são quem define as restrições a definir ao algoritmo. Restrições de como é que a informação é guardada, porque o algoritmo precisa ter dados de treino e para isso é preciso saber que a informação vai ser mantida mesmo que encriptada ou salvaguardada de alguma forma pela integridade, pelo tipo de informação. É preciso saber qual é que fica, qual é que não fica. Depois até de forma dependente da aplicação do modelo, que tipo de resposta ao modelo está disposto a dar ou não. Por exemplo, sobre um setor de saúde. Não é suposto que o modelo faça um diagnóstico muito pormenorizado, porque isso pode gerar alarmismo por exemplo, para utilizador final, portanto isso também depende, ou seja, até que ponto é que, ainda que o modelo seja capaz de prever determinadas informações, essa informação vai passar para o utilizador final. E depois o próprio treino do modelo, seja, toda a equipa técnica envolvida acaba por ter uma influência no treino. O Product Owner está sempre envolvido, mas o Product Owner à partida já tem algum conhecimento técnico, ou seja, já é uma pessoa muito orientada para o tipo de modelo ou do tipo de solução que se vai implementar. Normalmente eu diria que há sempre alguém mais acima, depende da posição que tem, pode ser alguém de uma comissão executiva, pode ser alguém que está mais ligado a interesses ao nível, mas não é bem funcional, se calhar em termos de legalidade, até onde é que vão permitir, em termos de atualização. Portanto Product Owner sempre, mas depois diria que essa pessoa pode ou não ser quem define determinados limites à sua solução, complementar. Normalmente há uma pessoa além do Product Owner.
5	Portanto, planeamento e design do algoritmo... data scientists, se calhar mais por dentro do projeto há sempre a opinião dos Business Users. Software developers e software architects para montar a pipeline e diria que era isto. Se calhar data analysts não tanto nesta fase, mas numa próxima de analisar resultados.
6	Pronto, aí entram os engenheiros de dados, os cientistas de dados, os data analysts também, porque depois temos que avaliar e avaliar as coisas. O project manager está incluído também.
7	For algorithm everything technical should be involved only involves the team of developers. They are the ones that they are hired for, so they should just; you can't give basically the project manager have the client needs and he has to explain those needs in not technical terms but just in human terms to the developers who will translate that to technical terms to for the machine to understand.
8	Ora será preciso um developer, que é o meu trabalho.Tenho um gestor. Neste caso ele é que toma as decisões. Mesmo que seja eu a ter contato com o cliente eu transfiro para ele e ele é que dá aprovação ou não.
9	Normalmente quem anda envolvido é o gestor de projeto, com o cliente, não é? Que faz o levantamento dos requisitos. Depois quando chega à parte do desenvolvimento, tirando aquelas partes mais técnicas, mas normalmente é o cliente com o gestor de projeto sempre que definem o que é que o software tem de funcionalidades e o que é que ele vai fazer. Eventualmente entra a equipa técnica. A gente só tem os programadores. Não temos, aliás, em Portugal, basicamente só temos programadores a trabalhar na área. Portanto, tens de ter sempre ali uma pessoa que faz isso, mas normalmente a implementação está muito, está muito feita só por programadores. Aquela parte científica do, está muito deixada de parte.
10	(Executive Management) Eu diria que diretamente eles não participam eles não participam. Quem está abaixo deles propõe um produto apresenta-lhes depois quem o desenvolve quem o cria quem o idealiza depois apresenta a sua solução ao diretor executivo com os trâmites legais com os trâmites técnicos e afins... o diretor executivo analisa, muitas vezes socorre-se da assessoria porque ele não vai conseguir analisar todos os projetos que as diversas equipas têm e lhe entregam. Recorre-se de assessores que validam o projeto e depois diz neste caso ao product manager ao product owner e à equipa que desenvolve o produto e à equipa que o idealiza, sim ou não porque no fim de contas as consequências finais são todas dele.
11	

Appendix B.4 – Answers provided for the Interview Question 4

	Interview Question 4: What are their specific responsibilities during the planning and design process?
1	Eu não sei como é que estão organizadas as entregas, mas ele tem de garantir que na data da entrega o produto está feito para ser entregue e com qualidade. Ao mesmo tempo tem de garantir que a equipa está equilibrada.
2	Pronto, o business lá está, eu acho que é um bocado a ponte entre o que é que é de negócio, o é que é o conhecimento de negócio para ajudar a equipa de desenvolvimento a interpretar os dados e a conseguir pronto, obter os resultados que se esperam. Depois, o data analyst, eu acho que é, são responsáveis por interpretar os dados e validar que, efetivamente, o algoritmo está correto e está a ter os dados que nós queremos. Os data scientists e engineers, pronto, para mim, como eu disse, acho que é um bocadinho a mesma coisa, pelo menos do que eu conheço. Têm a responsabilidade de recolher os dados e de os interpretar e de fazer o algoritmo. Depois, o software developer, eu acredito que seja a parte mais front-end e tecnicamente também a parte back-end, ou seja, que vai construir a interface para o utilizador depois usar, não sei qual é o contexto do projeto, mas seja lá em que contexto for, o mais provável é que o utilizador final tenha que interagir com a interface e eventualmente os developers vão ter de juntar esse algoritmo com depois a parte da interface. E depois o software arquiteto tem de garantir que tanto a interface como, no fundo ele o arquiteto do software. Os developers constroem, mas o arquiteto tem de organizar tudo para que fique tudo arquitetonicamente correto.
3	Então, é assim, o principal responsável do projeto há de ser o project owner, não é? Sobretudo, o que seja feito abaixo, ele vai ter que sempre dar a cara e aprovar. Mas o planeamento acaba por ter o papel de cada um, não? Porque cada um deles cá em baixo, da project team e depois do manager, cada um vai dizer mais ou menos o tempo que vai levar, ou aquilo que propõe, dar as suas métricas para chegar a... ao plano final. O software developer vai dizer quanto tempo se calhar é que precisa para desenvolver certas e determinadas coisas enquanto o arquiteto se calhar com o tempo vai dar ali várias opções de arquitetura, conforme o seu papel, a parte do que vão fazer é diferenciado. Pronto mas a nível do manager e owner vai ser também mais... Vão ter que lidar mais depois com a pressão do cliente e com os timings que estes tenham, (...). Estes ajustes depois já é mais a nível superior que vai ter que ser feito e pode sempre depois ser consultar novamente a equipa conforme as alterações sejam feitas, que tenham que ser feitas, mas depois essa discussão a nível de expectativas e diálogos depois há de ser sempre mais a nível superior. Porque há uma fase, tens a fase dos requisitos, não é? Que é levantado pelo cliente e depois há as sessões de discussão para que a equipa, o project team vá, entenda bem os requisitos que se quer, para depois poder propor, faça o documento laborativo do "to be", não é? Naquilo que vai ser delineado.
4	(Product Owner) No fundo é a pessoa que vai definir as características do produto a desenvolver. Ou seja, tal como o nome indica a pessoa que é quase encarregue pela definição dos requisitos daquela plataforma. Até ali o papel um bocado intermédio entre a equipa técnica e a equipa que se calhar não é equipa mas pronto, em termos assim do cliente acaba por ter aqui um papel intermédio e portanto depois dependendo do projeto acaba por ter vários papéis. Pode ser o próprio analista que é quem define requisitos diretamente para a equipa técnica, pode passá-los ao project manager ou a alguém analista da equipa técnica que faça esse trabalho, mas no fundo é a pessoa que vai definir eu quero que a plataforma faça isto, desta forma e normalmente é uma pessoa mais técnica porque eventuais limitações técnicas que possam existir têm que ser conversadas, lá está a conceção.
5	Sim, eu diria que isso é muito mais o trabalho de data scientist com a ajuda de systems software architect para desenvolver a pipeline e os data engineers também para o CSGTL e coisas desse género.
6	mas aí tem a ver com, são os engenheiros de dados, são os cientistas de dados que presumem que serão aqueles que estão a desenvolver um modelo de inteligência, a parametrizar os modelos.
7	But the project manager or owner should not tell them, the developers, what to do, how to do their work in details. Okay. Like no, you should use this, I want to have this at some ending point, add this, and have this result at the end and I want it like that, like that, like some kind of requirement. And they should do their work depending on those requirements. And that, but within the team, advising the team of developers, they should have like code review, algorithm review, so they can each one give the inputs to their inputs of each other's work. Code review or cross-testing maybe….
8	Ora primeiramente garantir que as coisas funcionam bem, depois que como pronto, aplicada a tua área no front-end tens de ter uma interface que seja fácil de utilizar por exemplo e que te tenha lá os dados que sejam mais interessantes de serem apresentados. Que o utilizador consiga chegar lá e veja, identifique aquilo que tem maior importância para ele. Depois também é aplicado a diferentes clientes e cada cliente tem a sua preocupação.se nós apresentarmos o software o software vá o SCADA e o cliente quiser alterações sim em contato com o cliente.
9	Depois quando chega à parte do desenvolvimento, tirando aquelas partes mais técnicas, mas normalmente é o cliente com o gestor de projeto sempre que definem o que é que o software tem de funcionalidades e o que é que ele vai fazer.
10	(Executive Management) E sabe e sabe de todas as alterações que houver pelo caminho. Caso a situação se altera obviamente se o projeto nasce A e termina A pronto tudo bem. Agora o problema é se ele nasce A e na verdade só se consegue terminar B e aí ele já não deu o aval ao B já houve mudanças, portanto teve que haver uma explicação das mudanças que houve e o diretor executivo provavelmente recorrerá a novamente dos seus advogados e dos seus assessores técnicos para novamente validar. Muitas vezes até esses assessores fazem parte da própria empresa que na verdade são outro project owner ao lado de outro produto que vem avaliar o produto do seu colega e juntam-se quase em equipa para fazer essa avaliação.
11	

Appendix B.5 – Answers provided for the Interview Question 5

	Interview Question 5: What are the responsibilities of the technical team during model development and testing?
1	Eu não sei a equipa validation, se quem é que faz validações, tipo os próprios conseguem fazer testes
2	Todos os que têm a ver com data tem que estar envolvidos a ajudar nos testes, a tirar dúvidas e a fazer correções. Os software developers também ao fim e ao cabo também vão estar a fazer testes também no software, portanto eles também têm que estar disponíveis igualmente. E aqui os business users acho que se calhar, talvez, vão ser as pessoas que vão estar mais encarregues de acompanhar os testes mais diretamente. E depois aí falam e... E pronto, e conversam com o resto da equipa para se arranjar soluções que se... que tiverem de arranjar, acho eu, diria.
3	E vão ter que fazer a primeira fase de testes. Portanto há de haver os testes de aceitação do lado do cliente. Dizem-te quais são os requisitos para que determinado entregado seja aceite. E eles têm que, quando entregam garantir que, pronto, aqueles requisitos são os mínimos cumpridos e depois passam para lá de cá e faz-se os restantes testes e garantir que aquilo está conforme e é fechado o requisito.
4	Desenvolvimento é sempre a equipa de desenvolvimento a ser o maior nível, porque estão envolvidos. Testes dependendo de quão grande seja a equipa pode haver uma equipa de testes determinada diretamente para isso. Se forem testes numa fase inicial, porque a partir de testes finais são o utilizador final também. Normalmente é a equipa QA, que a equipa que eles chamam Quality and Assurance, que é quem faz a validação suplementar, no fundo comunicam também só no intermédio. E pronto, depois o Product Manager pode haver ali mais pessoas, que é o Tech Lead da Equipa de Desenvolvimento, o Tech Lead da Equipa de QA, e depois dependendo de quantas equipas de desenvolvimento existem, pode haver mais tech leads. No fundo a Equipa de QA vai comunicar diretamente com o Product Owner e com o Project Manager, normalmente o Product Owner, (...). Mas pronto, no fundo vai alinhar, uma vez mais, quais é que são os requisitos da plataforma e validar o que é que foi implementado ou não, o que é que vale a mais ou não mas pode não ter visibilidade em termos técnicos de quanta informação por exemplo, está a agregar.
5	Sim, data scientist, software developer e se calhar também um pouco já data analyst, no modo de olhar para os resultados e ver se está até tudo certo, não sei. Data engineers para testar também. Se o pipeline também funciona.
6	A questão aqui é que a validação do modelo é uma validação técnica. Essa fase em termos de IA é uma validação técnica por isso não inclui os utilizadores de negócio é uma validação se o modelo se temos confiança ou não temos confiança nos dados que o modelo está a produzir, portanto é a mesma equipa. Ou seja a equipa de projeto e tem mais outra equipa que é a equipa do utilizador chave que são esses utilizadores chave que são os nossos key users que depois vamos usar para não só para o levantamento de requisitos, mas para a validação para o desenho da interface e para a validação de tudo. Antes de passar ao users tem de tudo ser validado e tem de ser feito testes e testes de usabilidade. Eles são dão-nos os requisitos e depois testam as interfaces que desenvolvemos. mas tipicamente há uma equipa de testes e há uma equipa que é responsável por fazer estes testes, sobretudo quando estamos a falar na parte do desenvolvimento da interface do utilizador é uma equipa diferente, é uma equipa... uma equipa de user experience e essa equipa de user experience é que tipicamente faz estes testes (...).
7	Nowadays you have the testers and QA. QA comes after you have delivered your featured project. It should make sure that it's as intended for the clients. Testing is more among the teams. I would say there's two levels. One among the teams where you cannot test your own code because it's not the right way. You should ask another member of the team to test your code, to test your feature. And then the project manager should test it to make sure that it was as he asked first. And after that, there should be a QA who just tries it as a client asks it. I think that there should be these three levels in the perfect world. In the perfect world.
8	Sim, até porque a forma como nós trabalhamos, não é, existe sempre partilhas de código então muitas vezes acabamos por usar aquilo que o nosso colega também fez e vice-versa. (Ok e vocês tipo registam os testes ou...) Não. Normalmente creio que sou eu ou quem estiver a trabalhar que vai acabar por resolver. Muitas vezes é impossível fazeres debug de tudo não é? Acaba por escapar e até ser detetado quem tiver na altura, é quem resolve.
9	Na nossa empresa é. É quem faz o programa, salta para o outro, fazem os testes e é feito assim. A segunda fase de teste é feita no terreno, verificar se está tudo a trabalhar, o cliente também ajuda nesse passo, não é? Mas não temos uma pessoa especificamente alocada. A fazer os testes de qualidade. É assim, depende da dimensão do projeto. Se for um projeto muito pequeno, não é? Acaba a ser tu a fazeres isso, acaba a ser só a pessoa a fazer isso. E muitas vezes temos uma pessoa que (...) é o comercial. O comercial também com o cliente, também faz esse tipo de verificações. Antes do sistema entrar em funcionamento, o comercial muitas vezes é aquele que faz o contacto. O primeiro contacto com o cliente e acaba depois por fazer também uma continuidade com o cliente. E depois os clientes normalmente também têm técnicos e também têm outras pessoas que também acabam por fazer isso, não é? Se contares com o nível do programador, tens ali uns três a quatro níveis de testes. Tens sempre no mínimo, no mínimo dos mínimos, umas quatro pessoas envolvidas. Quando está na fase de desenvolvimento, a parte dos testes e quem desenvolve é interno. Estamos a falar que os testes, uma parte deles são automáticos. São programados e são corridos automaticamente. Então há uma bateria de testes que se vão acumulando conforme se vai fazendo o desenvolvimento e o software vai se desenvolvendo e vai-se passando por essa bateria de testes para garantir que o software passa sempre esse conjunto de testes.
10	É isso é um teste primeiro com a equipa há feito pronto digamos o ensaio de qualidade na própria empresa o aval de qualidade dizer ok isto faz aquilo que se foi proposto e depois eu diria que há um teste em produção para o próprio cliente depois avaliar se depois realmente o software desenvolvido produz aquilo que foi proposto. E nesse teste em produção normalmente participa a entidade reguladora chamemos-lhe assim quem vai regular se realmente aquele software pode ser executado, participa também do ensaio em produção não dos ensaios chamemos-lhe em laboratório na própria empresa. (Developer) Faz testes ao próprio código e depois esse código também é entregue a uma pessoa que faz o trabalho dela é único e exclusivamente encontrar problemas no código que o outro fez, (...) uma equipa que só testes, mas lá está estamos a falar de empresas com dimensões bastante grandes que permitem ter este corpo técnico todo. Eu acredito que não seja a maioria das empresas portuguesas, aliás não acredito eu tenho mesmo a certeza que não é a esmagadora maioria das empresas portuguesas. Não tem simplesmente não tem a capacidade de ter esse corpo técnico todo.

11	Depende do caso do uso. Estamos a lidar com dados pessoais dados sensíveis que têm que estar protegidos sobre a RGPD, sim ou não? Se sim, tenho outra pergunta. O modelo é in-house, portanto não vai haver partilha de dados, nem partilha do próprio modelo, que possa ser utilizado ou testado por entidades externas? (Não.) Não, pois, então aqui a responsabilidade é do gestor de equipa de desenvolvimento e da própria equipa, portanto é do gestor mas penso que seja mais do gestor de equipa de desenvolvimento. Talvez, eu acho que falta aqui um cargo, software developers, falta software testers. Ou seja, uma coisa é desenvolver, outra coisa é testar. E as pessoas que testam têm que ser diferentes das que desenvolvem. E aqueles que desenvolvem e assim podemos ir ao Software Developers, têm que garantir que os testes que preparam para os testers de fazerem, passam todos os testes de usabilidade, viabilidade segurança. E têm que garantir que esses membros não violam os princípios de... fiabilidade, usabilidade. Tenho que garantir que o produto que estão a desenvolver está suficientemente desenvolvido para que seja submetida a testes. E isto tudo estamos a falar de testes internos também. Eu não sei qual é o âmbito, mas se estivermos a falar de uma coisa que possa ser vendida para um cliente externo, têm que ser primeiramente testadas in-house antes de poder apresentar os segundos possíveis testes para possíveis clientes. Portanto ainda faltam os testers aí, sim. Por exemplo, essa é muito boa, Quality Assurance, exato.

Appendix B.6 – Answers provided for the Interview Question 6

	Interview Question 6: Do you think there are roles for oversight or validation to ensure fairness and accountability during model development and testing? If yes, who assumes these roles?
1	Então a ideia do projeto veio de algum lado e é suposto haver um diálogo constante para garantir que os objetivos pretendidos se estão a atingir. Por isso não sei, eu diria que é suposto haver essas mensalmente, semanalmente, consoante os períodos de entregas, precisa haver uma comunicação constante ou com o product owner que estará em contacto com o cliente ou com o cliente em si ou se o cliente for dentro da própria empresa e garantir que as entregas estão alinhadas com os objetivos.
2	Acho que o project manager tem que supervisionar tudo. E é ele que vai definir quais são realmente os métodos de testagem e os, ai, como é que eu ia dizer? Pronto o schedule de tudo e como é que está organizado e como é que é feita a comunicação. Por isso, nesse sentido, o project manager também é muito importante. Mas, talvez haja uma entidade externa eu não tenho certeza do que estou a dizer, mas acredito que haja uma entidade externa que controle também estas testagens. Mas eu não sei bem o é que eu estou a dizer, só vou extrapolar, porque tratando-se de dados e tratando-se de inteligência artificial, acredito que uma entidade externa fosse necessária para validar realmente que está tudo conforme, mas não sei.
3	Sim. Ou seja, há até sempre uma responsabilidade, mas depende do tipo de... De metodologia de trabalho, se estamos a falar de uma coisa que é Scrum, ou se estamos a falar de uma coisa que é entregável ou pós-entregável ou se estamos a falar do outro, o antigo. Se estás a falar em waterfall, porque se estiveres waterfall, o que acontece é que tu tens os requisitos, eles produzem tudo e entregam, e aí não tens tanto controle, diria. Atualmente, o que acontece é que como tens equipas mistas em que acabas por ter pessoas do negócio e do cliente envolvidas seja nas daylies, seja onde for, acabam por ter sempre um ponto de situação. Normalmente do lado do cliente nunca é o... O gestor de projeto há de ser sempre alguém abaixo um técnico, (...) mas tem que ser alguém, se for um analista tem que ter um conhecimento técnico sempre, (...). (...) pelo menos uma ou duas pessoas que sejam responsáveis por isso. Pela estratégia de testes, implementem estratégia de testes porque quando tens um projeto é sempre aprovada uma estratégia de testes quais as fases que vai ocorrer, o que é que vai acontecer. Mas test qualquer coisa.
4	Sim. Sim. Product owner e Project manager. Depende aqui um bocado da constituição das equipes.
5	Pode ser alguém na mesma posição, mas mais sénior. Project managers também, de certeza. O project owner já numa se tiver alguma... se precisar de algo mais high level se calhar, não deve estar tão dentro dos detalhes, mas sim, também diria que sim mas não mais acima do que isso, se calhar.
6	se as coisas não estão a funcionar bem vai ter que falar com alguém que é responsável pelo desenvolvimento, com o project manager e depois tem que ser medido do ponto de vista de do ponto de vista de do sistema.
7	Sometimes, the project manager has the QA and sometimes you have to test your own code because the team is small. But in the perfect world, you should not test your own code because then you cannot trust how you, because as the developer built it himself, he knows how he should use it. We need someone who has not been building the feature or anything to build it as a new user. First. And then the project manager should use it as a client intended. And then the QA should test it.
8	(Project Manager) Sim, passa sempre tudo por ele. Existe alguém depois acima dele sim.
9	E depois acaba a ser supervisionado por gestor de projeto.
10	Eu diria que aí há uma partilha por duas pessoas o project manager e o project owner que acabam por ser os supervisores do resto do resto dos técnicos. Acho que são estas duas pessoas.
11	Sim, Project Manager, sem dúvida. Talvez o Project Owner e (...). Da boa execução do projeto, como todo. Então, se calhar eu não sei quem faz a curadoria dos dados, mas se for o Data Subjects ou a pessoa que recebe os dados do Data Subjects, também tem que garantir que existe uma curadoria. Ou seja, garantir se cumprem os padrões mínimos de qualidade, se possuir dados pessoais garantir que eles estão anonimizados, nesse sentido, diria que também teria envolvido um Data Protector Officer. Era preferível esse modo. Ou seja, primeiro equipa desenvolvedores, depois submeter ao QA e depois o QA valida ou manda para trás e só depois é que vamos... Tirar os testes com outro cliente ou cliente ou outro tipo de stakeholder, não interessa.

Appendix B.7 – Answers provided for the Interview Question 7

	Interview Question 7: Who do you consider that ensures that the AI system meets organizational requirements during its implementation?
1	Vai haver um período de testes os users têm de fazer testes para garantir que... não sei se a própria empresa tem tipo métricas que quer atingir e consegue fazer essas validações. Talvez este platform owner garantir que o serviço da plataforma está operacional diria.
2	Aí acho que já entra o teu quadrozinho mais abaixo, que são os executivos. E pronto, quem manda efetivamente na empresa e... Tu também tens aí o legal e ethics, faz todo o sentido. E aliás acho que ele também deve estar agora pensando bem, ele também deve estar envolvido no iniciozinho, onde tu me perguntaste quem é que era responsável pelos dados e por segurança dos dados. E ele também é muito importante. Eu acho que aqui é mais o cliente não é tanto a empresa que está a desenvolver portanto seria mais este quadro, para mim.
3	Eu diria pela experiência tenho, aquilo que é visível a nível dos requisitos, tens de ter sempre o gestor de projeto, esse é o que vai dar o carimbo final, a dizer que aquilo se é, mas quem vai estar envolvido nisso acaba por ser também um bocadinho a nível de... tu para dar esse carinho final vais ter que ter sempre a garantia que os testes correram bem, que os testes de aceitação estão cumpridos, que essa fase foi toda passada e que agora quando testaste e fizeste as últimas análises que aquilo estava tudo ok. Aqui até podes envolver o utilizador final, diria. E depois, a nível de direção. Não sei se podes envolver também nessa fase ou não. Eu acho que pode ser sempre envolvido, mas se faz sentido ou não depende do tipo de produto.
4	Normalmente project manager e aquela pessoa que eu falei que pode ser o product owner, mas que muitas das vezes não é, no fundo é quem decide, quem tem a responsabilidade para prestar comissão executiva ao equivalente da empresa, que contrata, de perceber o que é que foi implementado e como, agora não tenho certeza. Depois o executivo, talvez também, dependendo, eu diria que não faz tudo, provavelmente parte das pessoas que estão envolvidas naquele retângulo interior, do control, o model, o maintenance, provavelmente também tem alguma. (Platform Owner) Então neste caso acho que esta pessoa também está. Imagina eu também não consigo excluir os end users, porque das coisas que se fazem sempre em AI, por muito pouca responsabilidade que o end user tem, é a aceitação de termos e regulamentações de uma determinada plataforma.
5	Sim, devia ser um bocado por parte das duas equipas se calhar a equipa que desenvolveu, confirmar que ficou tudo como tinham descrito inicialmente e a equipa que que vai integrar. Sim. Também ter testes que confirmem que dá tudo como era esperado.
6	Eu acho que aí vai a equipa ou seja, os analistas não ou seja, a equipa tem que ir para lá ou seja, tem que ir tem que o modelo, todo o sistema tem que correr no ambiente tecnológico do cliente e portanto aqui entra obviamente o project manager...
7	I think after everything has been tested and validated, then the product manager or the project owner, yeah, should be the project owner who says, 'Okay, everything's validated. We can now implement that on production to deliver that to the clients.' Okay. Everything has been validated, so okay, let's go. That means that work has been done, so the developer will not be involved at this stage anymore because everything has been validated. So, okay, I can send that to the clients and should be shipped to them. In my opinion, it's a technical part. So once the QA has been validated, we think, we implement it, and then the client will receive what they need and they start to use it.
8	
9	
10	
11	

Appendix B.8 – Answers provided for the Interview Question 8

	Interview Question 8: What are the responsibilities of the Platform Owners or the Operational teams?
1	Eles têm de garantir sempre que o serviço está a funcionar independentemente de terem sido eles a criá-lo ou não. Podem manter um contacto mais ativo com quem criou ou não, então podem assumir eles, podem continuar a desenvolver ou erros ou whatever. Opá, eu acho que o cliente vai ter sempre de ter uma equipa tipo a equipa backup que vai sempre ter de garantir que os serviços estão operacionais. Porque ao fim de contas é o cliente que dá a cara e os users depois não vão saber se foi a própria empresa que criou ou não algum serviço, mas têm garantir que o mantém pelo menos. Por isso, nesse sentido eu acho que eles têm alguma responsabilidade em pelo menos manter os serviços que eles oferecem ativos e a funcionarem bem. Depois a maneira como o fazem acho que acaba por ser depois um bocadinho mais gestão interna.
2	O legal e o ethics é o que eu disse há bocado. Eles são essenciais. Executive management é pronto, o executivo de management tem responsabilidade por tudo. Lá está... tanto eles como o executivo da equipa de desenvolvimento ou da empresa de desenvolvimento são os dois muito importantes. Acho que a reputação tanto do cliente como do... Da empresa que a desenvolver, acho que estão ambos em jogo, mas acho que fica sempre um bocadinho mais do lado do cliente, porque é o cliente que dá a cara do produto, porque o produto é do cliente. Depois, model maintainer... seria talvez uma equipa funcional dentro do cliente que está encarregue. Será? É isso? Ok. Ok, então nesse caso é o que eu digo que eu falo mais com a equipa funcional do banco. É como se fosse essa entidade, ok. Eu acho que eles são muito importantes na definição dos requisitos, na testagem e depois, lá está, como o próprio nome diz, na manutenção. E ver se as coisas estão a evoluir bem e pronto e chatear quem tiverem de chatear.
3	Na implementação vai ser sempre a supervisão. Tenho que garantir que aquilo que está sendo desenvolvido é aquilo que ele pediu e está a corresponder ao esperado. Por isso é que depois há a parte do levantamento dos requisitos essa é muito importante. E aquela fase inicial em que há as conversas entre uma parte e outra para definir aquilo bem, depois, quando chegarmos a esta fase, não tínhamos coisas desenvolvidas que depois não está nada a fazer o suposto.
4	-
5	Sim, a responsabilidade em termos dos dados que estão a usar se estão se estão estão estão de acordo com as regulações todas que existem, não estão a usar dados pessoais que não podem, usar informação pessoal. Por exemplos, os algoritmos também há sempre considerações se calhar de fairness e equidade depende um bocado do algoritmo. Isso também é importante hum sim, diria que essas são as duas coisas bem mais à cabeça.
6	
7	And from that, if there are any kind of bugs, they should report to the project owner in order for him to give the developer the task to fix. The bug on his thing that is not needed, any bug or any conflict or any bad behavior.
8	(vocês ficam sempre em manutenção do programa ou o cliente acaba por ter a manutenção do programa?) Não, sempre nós.
9	O cliente há de ter a utilização normal da aplicação. Há de detetar um problema qualquer. Às vezes não é um problema de desenho. Pode ser uma pequena modificação, um comportamento que não se estava à espera. E depois ele volta a entrar, manda para o comercial, não é? A informar esse facto. O comercial manda para o gestor desse projeto. E depois o gestor desse projeto vai entrar outra vez na parte do desenvolvimento. E no desenvolvimento é sempre feito aquele ciclo. (ciclo está na resposta sobre o desenvolvimento do programa) (...) quase de certeza, vai aparecer algum problema, porque as coisas estão sempre a evoluir, o próprio negócio está a evoluir. Os nossos sistemas nunca são independentes do resto, nem do hardware, nem do software onde ele está instalado e vai haver, ocasionalmente, alguma coisa que vai fazer. Ou o Windows faz uma atualização ou qualquer outra coisa e por aí aquilo vai tudo. Mas, pronto, essa parte depois segue sempre para esse ciclo, estás a ver?
10	
11	

Appendix B.9 – Answers provided for the Interview Question 9

Interview Question 9: What mechanisms have you seen for collaboration among all stakeholders mentioned during the implementation and integration phase?	
1	
2	Sim, acredito que haja hierarquias pelo menos, é o que eu tenho visto nos meus projetos da TI e acredito que no de AI não seja muito diferente. Por exemplo, a equipa de desenvolvimento pode ou não ter contato direto com o cliente, mas tendo, deverá ser sempre aí com a equipa do module maintainers e o data controller, eu não tenho certeza do que é isso quer dizer, neste contexto, mas partindo do princípio que é como se fosse um profissional de data mas do cliente. Pronto partindo desse princípio, acho que acho que até deve passar primeiro sempre pelo data e depois é que se calhar vai ao funcional, acredito eu. E quem é que fala com o cliente diretamente? Isso depois vai depender sempre do projeto, há projetos que são um bocado mais formais e tens que abrir um ticket e depois eventualmente marcar uma reunião com os intervenientes. Há assuntos que se calhar vão ter de passar para cima e ir para o executive management ou não. Normalmente nos projetos que eu conheço o executive management, costuma-se ter reuniões mais periódicas com eles para dizer como é que está a correr o projeto etc, mas o dia-a-dia é mais com o funcional e com o data controller, pronto. E, pronto, agora como é está organizado vai sempre depender, mas a ligação deve ser minimamente direta. Primeiro tens de ter uma plataforma, seja ela o Teams ou o Skype ou seja o que for, e se possível o presencial, não é? Depois tens de ter a organização do... não sei se o projeto é por sprints, mas imaginando que é por sprints ou mesmo que não seja por sprints, tens de ter uma plataforma para organizar, não é? Que é gerida pelo project manager. Isto pode ser o Trello, pode ser o que tu quiseres. Depois tens de ter a plataforma onde vais interagir com o cliente. Isso tanto pode ser por e-mail que é um bocadinho menos prático ou pode ser uma plataforma específica para isso. E acho que é só.
3	A parte dos testes é importante, terem ali o processo todo passado e depois tens, podes fazer a questão das apresentações, aquela coisa dos "pocos" dos conceitos em que podes garantir, mas isto mais a nível superior, para demonstrares que realmente aquilo está ok. Depois tens os comitês tens que ir lá apresentar as evoluções e o estado da situação, portanto aí também é uma parte que é um processo de integração tens ali as partes interessadas, mais a nível da administração, direção, tens o responsável também do fornecedor, portanto estão ali todos. É claro que não falam só daquele projeto, falam dos vários que existem, mas quando falam a base do projeto o que esteja a fazer é preciso dar o feedback, dizer como é que está, que é que está a ser cumprido ou não e dar a garantia dos requisitos também.
4	Por isso dizia que normalmente o intuito é que haja a colaboração porque ambas as partes querem tirar proveito da relação que existe, do desenvolvimento da ferramenta com os pré-requisitos e as necessidades que o cliente pede, ao mesmo tempo que precisa de haver sempre colaboração. Mas diria que a maior responsabilidade é sempre da empresa que contrata.
5	Sim sei que existe colaboração, nunca tive essa experiência mas não sei, não consigo dizer mecanismos em específico.
6	Uma vez mais, o que falta aqui tem a ver com a metodologia de desenvolvimento do software isto é um software e depende da metodologia os mecanismos de articulação depende da metodologia que é seguida e do processo de desenvolvimento. Existem várias metodologias de desenvolvimento de software ou é um processo de desenvolvimento em cascata ou um processo com metodologias ágeis e tudo isso depende daquilo que foi acordado pelo cliente cada uma destas metodologias de desenvolvimento de software tem os seus mecanismos e as suas formas previstas de interação entre o cliente e o fornecedor. ... em cascata é mais waterfall, portanto quer dizer que inicialmente há várias reuniões com o cliente e depois vai desenvolvendo ou seja há menos interação com o cliente. Agile há todo depende da definição é feito por sprints há várias reuniões há um acompanhamento que pode ser um acompanhamento diário e depois há reuniões de sprint de entregas de sprint com a equipa mais alargada e depois tudo isto é feito de uma forma com uma colaboração muito mais estreita. Uma vez mais depende das organizações depende da maturidade depende do que é que é do que é que estamos a falar de inteligência artificial também.
7	They should have some kind of, they should be in contact with some kind of ticketing, you know, or something, or email, or I don't know, any sort of software, where the user, the client is using the products. And when they want something that is not producing what they need, they should just send the tickets, where the product owner should just take that into account and report that to the developer. Give the task to one of the developers who is concerned and ask them to fix it. And there we go, the process, again, or new feature, testing, approve, QA, and ship it back to the, ship it back to the, to the clients.
8	(são notificados têm tipo um programa específico com tickets ou...) Temos. É 50-50 nós gostaríamos de pôr-se tudo tratado por e-mail mas muitas vezes liga pronto. (reúne ... semanalmente?) Pode-se dizer que sim vá
9	
10	Na realidade não é isso que acontece porque o tempo urge e nem sempre é possível a presença de todos nem sempre é possível a disponibilidade, (...) que as reuniões acabam por ser o entrave à produção porque as reuniões alongam-se as reuniões agendadas para 15 minutos à sexta-feira de manhã na verdade transformam-se numa reunião de sexta-feira de manhã que preenche a manhã toda e coisas assim do género. E porque normalmente o assunto que poderia ser tratado via email rapidamente é arrastado depois para a reunião, porque aproveita-se não é. Isto é o que acontece na grande maioria das empresas de hoje em dia e é uma queixa, eu acho que transversal, aos engenheiros informáticos, aos engenheiros eletrotécnicos e até aos web developers e web designers. E acredito que os data analysts e aos data scientists aconteça o mesmo. Por acaso tenho colegas que o são, data scientists e data analysts e confesso que ainda não falei com eles sobre o assunto, mas acredito que seja transversal.
11	

Appendix B.10 – Answers provided for the Interview Question 10

	Interview Question 10: Who is responsible for monitoring the AI system's performance and solving issues like bias, errors, or failures? For example, if a bias is detected, who do you think should handle it, and why?
1	Eu diria aí uma equipa uma equipa que ficasse responsável por manter o serviço. Mesmo mais focada nisso.
2	Por um lado o cliente o data controller, essencialmente vai ter que estar atento, mas por outro lado o data controller da equipa o data controller e toda a equipa de data também tem que estar envolvida.
3	Sim, depende do contrato Durante o período de garantia, tem que ser o fornecedor a custo zero. Após esse período de garantia e também, lá está, depois de dar o seu, imagina, houve o desenvolvimento aprovaste e depois vais dizer que afinal não está a trabalhar em condições a responsabilidade já é tua. O que é que acontece? Hás de ter uma equipa de manutenção, à posteriori e esse valor depois será cobrado pela manutenção. Terás de pagar novamente para resolver um problema que se calhar a responsabilidade devia ter sido a anterior mas como só deste conta, nesta fase de manutenção, à posteriori da entrega (...). Mas há de ser a equipa que seja responsável pela manutenção. Já não vais imputar responsabilidade ao que produziu, porque já não podes porque já entregou o produto, o produto foi entregue a partir daí é teu.
4	Normalmente é suposto que sim, ou seja, por exemplo, uma equipa técnica, à partida nunca terá a mesma responsabilidade. Por exemplo, a KPMG é contratada por uma empresa, um banco. Se eu houver tipo de informação, dados que são demasiado guardados que se calhar não deviam ser, eu não digo que a empresa que desenvolveu a ferramenta não seja imputada por qualquer culpa, mas diria que a empresa responsável será sempre... a que pediu a ferramenta até porque isso normalmente é salvaguardado em contratos. (...) houver por exemplo, um período de manutenção ou um contrato seguinte de manutenção. Diria que continua a estar envolvida a equipa técnica, ainda que se calhar numa dimensão muito inferior. Há alguém do cliente porque normalmente este tipo, não sei exatamente quem é o cliente, mas deve haver talvez o platform owner que esteja envolvido em esta fase. Que é quem fará o controle se a plataforma está a funcionar de acordo com o previsto. Se não, sobre uma coisa ao nível interno, ou seja, a equipa de manutenção está ao nível do cliente, diria que só o product owner e as pessoas envolvidas nesse processo de manutenção.
5	(...), normalmente tens uma equipa que está encarregue de avaliar ativamente o modelo que está em produção ou algo semelhante e que tenha formas de ir avaliando ou detetando anomalias para depois conseguirem corrigir. Portanto se fosse por exemplo que o modelo estava biased aí tinha de ser data scientist para, ou tornar o modelo de forma diferente ou podias pronto, outras técnicas de rebalancing ou treino dos dados ou algo semelhante. Eu estava a pensar se podia ser também parte dos data engineer nisso não, porque o calha sempre nos data scientist hum pronto, seria sempre data scientist porque seria nesse caso, ou usar os dados de forma diferente no treino ou tornar o modelo de forma diferente, para terem atenção.
6	Estes modelos não correm sem manutenção, ou seja, e sem treinamento o que quer dizer que tem que haver uma equipa. Ou seja, tem que haver uma equipa de cientistas de dados de engenheiros de tudo isto que tem que ser tudo aquilo que está ali do project team tem que estar do lado de lá e a equipa tem que estar em articulação porque isto não é treinado uma vez e fica lá e esse é que é o problema. Isto nós construímos, treinamos mas a organização tem que ter competências para depois re-treinar o modelo continuar a trabalhar com os dados que existem ou seja, tem que aprender e tem que fazer. Não sei se é a equipa toda novamente, não precisará provavelmente de um arquiteto de software. Lá está tem que perceber mas tem que haver do lado da organização cliente a capacidade de retreinar os modelos de afinar de manter e esse é que é o problema da sustentabilidade deste tipo de projetos sobre um prazo. É que as pessoas acham que isto é uma coisa que se compra é um computador que se compra fica lá e é ligar à corrente e não é. Ou seja o sistema tem que ser mantido, o sistema tem que ser porque tem que ser retreinado porque o contexto muda e portanto tem que haver competências do lado das organizações.
7	So, if that's the case, you should have some helpdesk or some provider who should be monitoring the servers and the, you know, the performance of the server, or anything. In case of performance crash or website crash.There's another team. And there's one monitoring the system, the network, the servers, uh, the performance, the memory, and everything. It depends. It depends on who the client is. If the client is a company, he hired an external team of developers, he may have to hire another team for the, uh, for the hosting, for the hosting parts. And, uh, sometimes, if the client is just an individual user, for example, for, like, let's say, his Amazon. He should have his own team of developers and hosting and everything.
8	Quando se tratam assim de erros mais graves, muitas vezes sim ele pronto acabamos também por informar-lhe. Quando são coisas mínimas que não têm assim extrema importância não é muitas vezes pronto são coisas mesmo... aí não. O cliente crê que olha como a empresa sendo responsável, não a pessoa x, a pessoa... pelo menos pelo que eu tenho visualizado até agora não é experienciado, normalmente o cliente não procura uma pessoa em particular. Isto teria que ser a empresa não é que a fim ao cabo... deveriaa existir um elemento que pronto que controlasse e que desse o como é que é dizer a aprovação final e então pronto por muito que alguém cometa um erro seja intencional ou não depois aquilo segue para alguém vá haver uma entidade que regule neste caso eu tenho a pessoa responsável por mim. E sem querer não vou dizer que ele é culpado por ter aprovado, mas, passa um pouco por aí. Pronto é a parte de teres responsabilidade num sítio não é.
9	Passa sempre pelo gestor de projeto. Ok. O ciclo é quando colocas a aplicação no cliente. O cliente começa a mexer na aplicação. Deteta alguma anomalia, vai informar o comercial dessa anomalia. O comercial vai informar o gestor de projeto. O gestor de projeto vai entregar ao responsável da equipa da programação. E o responsável da equipa da programação depois já dá a entregar à pessoa que vai acabar por mexer nisso. Vamos falar assim, houve uma má interpretação da equipa de desenvolvimento dos requisitos do cliente. Quando chega o cliente, o cliente vai identificar aquilo como se fosse um bug. É mesmo por causa desse tipo de coisas que tens a parte comercial e o gestor de projeto antes de chegar a equipa técnica. Porque há sempre uma barreira linguística, não é? Porque os técnicos têm uma maneira de falar muito específica e os clientes já não.
10	eles não permitem o início da produção sem que o software e o hardware que foi lá aplicado cumpram os trâmites mínimos. Eles enviam o parecer deles depois da avaliação e cabe depois à empresa que está a desenvolver o produto corrigir e eles depois fazem uma nova auditoria e se tiver corrigido aí sim certificam e dizem ok podem começar a produção. Há o chamado fiscal que é a pessoa que vai que é a pessoa responsável pela fiscalização que por si depois também tem um grupo de técnicos cada um especializado na sua área. Um vai ver se o software está ok o outro vai ver se o hardware está ok, partilha os dados partilha os dados com o superior e o superior comunica à empresa o

	resultado dos relatórios. Normalmente o project manager, ele depois recebe analisa vê em que área é que se enquadra neste caso se estiver tudo ok pronto está ok está arrumado. O problema é quando há lacunas e aí ele tem que encaminhar para devido membro da equipa a resolução da lacuna.
11	então seria um grupo de pessoas alocadas à monitorização e manutenção e suporte, sim sim. São estas as pessoas que resolvem quando o sistema tem problemas de envisiamento erros ou falhas exatamente se foram eles que desenvolveram, são eles que têm que dar fixo a qualquer coisa. No entanto, novamente aqui apelando à responsabilidade, quem desenha sistemas e quem os concebe sabe perfeitamente o que está a fazer. Estas máquinas não vão ganhar vida por si próprias a não ser que nós queiramos que ela se reinventa ela própria e para isso ela tem que ter alguma instrução para tal. Se existe um enviesamento das duas uma, se isto é criado in-house, treinado por nós e foi afinado o fine tuning por nós, nós empresas ou desenvolvedores, tem que ser equipa de programação gestor e todos os envolventes que são responsáveis por levar a bom porto esses tais enviesamentos. Se eles estão a acontecer é porque não os preveram ou então não o corrigiram.

Appendix B.11 – Answers provided for the Interview Question 11

Interview Question 11: Is the End User responsible for the outcome of the AI system?	
1	Eu acho que ele tem alguma responsabilidade tendo em conta os inputs que dá. Se ele pode tomar uma série de decisões e se ele as tomar de maneira diferente o outcome vai ser diferente. A não ser que o algoritmo queira sempre dar o mesmo resultado.
2	Responsável não mas é imprescindível até porque normalmente acho que estes algoritmos vão também aprendendo depois à medida que o end user está a utilizar. Eles são essenciais, mas não diria responsável. Gostava que fossem, que as pessoas fizessem a sua parte e tivessem responsabilidade, mas acho que não se pode obrigar as pessoas a terem essa responsabilidade.
3	Sim, mas é responsável porque ele é que dá, se for alguma coisa de inteligência artificial... Nesse sentido, agora a pensar um bocadinho sobre isso, parece-me que ele seja responsável, porque, isto é, como uma receita, não é? Dependendo dos ingredientes que lá colocas aquilo pode ter vários resultados. Porque aquilo acaba por ter uma vida própria, esta é a diferença entre uma coisa que é programada, tu sabes, tu defines muito bem o requisito, muito bem os caminhos a seguir, o fluxo muito bem conhecido e portanto aquilo quando dá o mais um vai dar o dois, nunca dá dois e meio, portanto aquilo tem uma resposta sempre certa neste caso, faz.
4	Portanto eu, enquanto utilizadora por exemplo, da ferramenta do OpenAI, eu tenho que saber que a informação que lá ponho não é confidencial. Portanto de certa forma responsabilidade para aquilo que eu lá faço também é minha. Se calhar menos do que o platform owner e os outros envolvidos, mas sempre algo, portanto estes quatro eu conheço muito. Sim, porque isso mesmo ao nível da empresa, por exemplo, informações internas já é considerado derivado do avanço do chat GPT, não é proibido usar, mas há muitas empresas que gostam de fazer chat GPT interno embora obviamente nunca tanto poderoso como o original, porque a equipa não está focada no desenvolvimento de uma plataforma equivalente, era evitar fugas de informação. Agora, o trabalhador, ainda que o colaborador ou qualquer pessoa que utiliza uma plataforma, cujo algoritmo não pertence à empresa, acaba por fazer ali um leek de informação, ou seja, faz, ou tem a responsabilidade de fazer a encriptação de informação, que é por exemplo, determinadas análises Excel, substituir todos os utilizadores de informação sensível por letras ou nomes ou o que seja, ou então simplesmente não use, porque se acontecer algum tipo de leeking em dados, derivar desse tipo de informação, a culpa recai sobre o colaborador, porque esse tipo de responsabilidades que ele tem.
5	Quer dizer, se estou a pensar num exemplo, tipo, imagina lá as language models. O end-user é responsável se usar de forma não benéfica. Estou a pensar nesse caso. Portanto aí, sim. Sim, é isso que eu estava a dizer se for algo que pronto, que vem mesmo do programa ou do algoritmo e não é um outcome desejável, não foi o user e tentar buscar ou usar de forma não foi suposto, se calhar, aí não tem responsabilidade.
6	Não, não os end users depende daquilo que, ou seja, eu diria que não. Ou seja, há um conjunto de regras que eu vou usar. por exemplo de um modelo preditivo de concessão de crédito não será certamente a pessoa o funcionário do banco que está a usar o modelo que é responsável por uma decisão errada do modelo. O end user aqui estou a ver como seja o funcionário, não é do banco que está à frente com o seu cliente o gestor de conta que vai dizer se... o gestor de conta não, o gestor pronto... que está lá que vai dizer se conceda ou não concede crédito é ele que é responsável? Ele tem que dar as informações todas para o modelo mas depois o modelo toma uma decisão e depois a seguir a decisão é obviamente do gestor que tem que ter. Mas tipicamente não é só dele, de quem é que é? É da parte da gestão exatamente que tem aqui executive management, ou seja, a gestão é que é responsável no limite por aquela decisão porque se nós estamos a dar uma ferramenta para os utilizadores da organização usarem há uma responsabilidade formal é também é da gestão de topo.
7	It depends on what kind of user, what kind of program. I think it can vary. If it's just a chatbot, no, because it's just talking, using it and talking to the chats and that's all. But if what he uses is impacting someone else directly, then that means that he just uses the project in some intended way. And that should be the end user's responsibility. Let's say for example on the web.... On the, on the page that you can upload, you can upload file. You will be using that to upload file and send it like in a document, important document to, to someone. What if someone uploaded a virus to send it to someone? Because there the responsibility would be on the, on the end user because intentionally he's uploaded something that can affect, hurt someone else. For ChatGPT, you get data, you get a result depending on the prompt you give it. So I think that maybe it's important, it's how you view that data and how you use it, that's, that is important, and not the data itself.
8	Eu creio que sim sabes porque pronto aquilo tem falhas tem muitas falhas como tu também deves saber. E cabe a nós não é saber interpretar o que está lá. Sim, se fui eu que inseri a informação. Creio que a partir do momento em que ele me debita qualquer coisa e que eu estou a utilizar, eu sou responsável não é. Eu escolhi usar essa ferramenta.
9	O end user tem influência. Se tu pensares num sistema simples, que a gente já está usando agora, como o ChatGPT. O end user é o responsável por fazer as prompts. Se ele fizer alguma coisa que o sistema não está à espera. E não obter a resposta que ele está à espera. Pois a responsabilidade é dele. Há sempre uma responsabilidade do end user. Mas no caso da inteligência artificial, e é este o problema, é que quem desenvolve não tem conhecimento total do comportamento da inteligência artificial. Eu imagino uma coisa que as pessoas às vezes não têm noção da responsabilidade quando colocam as coisas no chat GPT.
10	E relativamente a essa eu acho que o utilizador final invariavelmente terá sempre uma cota parte de responsabilidade é quase se... há uma coisa que corre mal é que o utilizador... é que depois há dois utilizadores, há dois utilizadores finais. Dando o exemplo de dos carros da Tesla que eles têm inteligência artificial para conduzirem sozinhos, mas o utilizador o utilizador final é logo à partida avisado de que cuidado que esta inteligência artificial tem as suas limitações e pode provocar um acidente. Ele está consciente ele foi avisado. E se houver um acidente não é a Tesla que é responsabilizada é ele. E ele depois inicia um processo legal contra a Tesla porque lhes vendem um produto que de certa forma tem lacunas, mas na verdade aos olhos da lei ele foi o responsável. Sim, ele o terá sempre sempre uma cota. É impossível fugir à responsabilidade do utilizador final eu acho que é eu acho que é por um e nós até podemos ver e observar que o utilizador final não teve culpa ele fez tudo bem mas não dá foi ele que utilizou foi ele que utilizou (...). E só poderá só deixará de acontecer quando não é possível existir o utilizador final e a inteligência artificial se começar a usar a si própria e aí não há um utilizador outra inteligência que utilizou uma inteligência, acho que é a única maneira do utilizador final não ser não ter pelo menos um pouco de responsabilidade.

11	claro. Também, não é? Se aquele sistema de inteligência artificial foi desenhado novamente para cumprir uma tarefa e o utilizador não está a ir de acordo com a tarefa com que ele foi desenhado, é claro que é responsável ok parte boa responsável sim.

Appendix B.12 – Answers provided for the Interview Question 12

Interview Question 12: What are the responsibilities of external stakeholders (e.g., regulators, lawyers) during the Consequences phase?	
1	
2	As auditorias têm uma responsabilidade muito grande, não é? São eles que têm que avaliar que as coisas estão a correr bem e nem é só em projetos da IT, em tudo. Opinião pública, não acho que... enfim acho que é só uma opinião pública. Não acho que tenham responsabilidade. Acho que influenciam, mas não têm responsabilidade. Data subject... sim, os data subject... Eu acho que a certo ponto, sendo sincera, eu acho que os data subject não têm responsabilidade nenhuma e a única coisa que eles vão ter é, se calhar interesse. Independent evaluators, não sei o que é que são os avaliadores independentes não faço ideia. Porque eu diria só a auditoria mesmo. Agora, é claro que um advogado pode intervir, mas à partida porque é que um advogado deveria intervir, se ele for em nome de alguém? E isso quer dizer que as coisas estão a correr muito mal (...), portanto não acho que isso seja uma responsabilidade. Quanto muito acredito que o governo por exemplo possa ser um bocadinho stakeholder nisto ou ter uma responsabilidade.
3	Eu poria as responsabilidades mais a nível de auditoria e fiscalização, ou seja, normalmente os utilizadores é que estão pelos problemas, quem o utiliza à partida é que vai sentido os problemas (...). Ou seja, o utilizador final é que sabe a necessidade com que utiliza. Será ele sempre aquele que vai poder dar o melhor, fazer a melhor análise ao produto.
4	A que desenvolveu, eu acho que poderá sempre receber essa informação, provavelmente também poderá procurar de forma autónoma e no começo à partida será a empresa, ou seja, este trabalho normalmente quando se contrata uma empresa para desenvolvimento depende um bocado de quantas equipas estão envolvidas. Se for só uma equipa de desenvolvimento à partida não tem muita responsabilidade sobre a informação que é passada ou seja, tem alguma, tem que saber o mínimo do que é que é permitido ou não, mas diria que se for uma coisa muito específica é a responsabilidade da empresa ou da plataforma é fazer esclarecer e chegar essa informação para a empresa contratada. Às vezes há projetos em que a equipa maior é do lado do cliente e outras vezes é o oposto, por isso depende.
5	Courts and regulators sim. Por definir regras que impactam o desenvolvimento do programa. Não só o desenvolvimento, mas também ... o desenvolvimento e definem a forma como pode ser usado e não pode ser usado. (...) por exemplo, vejo agora muito a parte dos courts and regulators cada vez mais têm de regular novas tecnologias e modelos novos que estão a sair e, portanto, aí muda um pouco a responsabilidade.
6	
7	They should create the law to, you know, make, to protect, uh, people of, uh, anything you want, anything anyone so that the AI should not be used in any wrong way. Let's say that you have AI, you know to generate pictures. You know, or to, you know, anything that relates to deepfake, for example, you should have a law for that, to prevent people from using that kind of technology in the wrong way that would affect another human being, an artist whose work would be stolen. Basically, AI just taking his impression is the other person's work, and just rebuilding in his own style, or involving another person because his identity will be used in this picture. So yeah, you should create law to just prevent not the AI, but the people for doing that, for using the AI the wrong way. The feedback from them matters. It would be really important for them to, it would be really important for them to understand the consequences of the AI. And to, I don't want to say to pre-shots what will be is, but to react quite quickly about anything bad that had happened because of AI. Because it's really something that is evolving really fast. So, if they take too long to create laws or anything, or to react to that, a lot of mess can happen. It's depending on each country, yeah. You cannot create a law for everything. Developers can have good practice, but for law, it depends on the country.
8	
9	(Já aconteceu alguma vez o cliente pôr um advogado em causa ou assim?) Se uma coisa destas ocorrer durante o projeto, pois normalmente há só duas saídas possíveis. Um impasse e é resolvido sem chegar a essas vias judiciais. Pode haver a ameaça primeiramente daquilo acontecer. Ou então o projeto acaba. É logo terminado. Porque há uma relação de, com os clientes há sempre uma relação próxima. E os projetos normalmente só se conseguem desenvolver quando o cliente colabora. Quando o cliente deixa de colaborar, pronto, aquilo vai logo pelo charco. Mas é uma coisa um pouco específica da nossa área. Porque a gente somos um pouco alfaiates. E todos os produtos que a gente faz são feitos à medida especificamente para aquele cliente. Então sofremos um pouco disso. Mas se for um projeto de uma empresa grande, que é um projeto para vários clientes. Normalmente os clientes não podem dizer nada. Podem se queixar de uma coisa não funcionar ou outra. Mas nunca podem fazer isso.
10	(auditoria externa) Acontecerá quando se está a trabalhar para uma indústria bastante exigente. Vou-lhe dar o exemplo da indústria farmacêutica que a indústria farmacêutica sim avalia depois o software que lá é usado com as limitações técnicas que o próprio avaliador tem, porque para o avaliador conseguir avaliar tudo ao detalhe precisa também ter um corpo um corpo técnico bastante bastante competente e a maioria das vezes não o é, porque temos as equipas de desenvolvimento tem o corpo técnico mais competente depois os avaliadores acabam sempre por não conseguir chegar ao mesmo nível. A avaliação acaba por ser um pouco mais superficial mas a indústria farmacêutica avalia não se mete software nenhum por exemplo na OVION sem que vá lá o infarmed verificar se o software cumpre os requisitos mínimos.
11	há um sistema auditorias sim sim Por acaso isto vai um bocadinho ao encontro daquilo que diz o AI Act. É uma coisa que, para nós que estamos aqui na União Europeia, vem limitar bastante os sistemas AI para os sistemas privados e públicos em que as únicas exceções é apenas a investigação e a academia. Portanto, eu penso que segundo o AI Act, tem que haver pelo menos... eu já não me recordo já soube disto... auditoria atenção, porque todas as outras coisas que o AI Act diz sobre o modo como utilizamos, que tipo de dados utilizamos, para que finalidades, isso está tudo já... Regulamentado, agora a nível da auditoria não sei por acaso excelente questão... por exemplo, imagine que uma empresa desenvolve um sistema que é todo bonito portanto ele cumpre a tarefa mas que por trás tem algo malicioso que apenas a empresa ou o cliente ou empresa que conhece isso tem que ser tudo revisto. E quando estamos a falar de inteligência artificial cada vez faz mais sentido esses sistemas estarem sob constante auditoria, sob constante prova

	do próprio sistema. Portanto sim, certamente vai haver entidades certamente vai haver necessidade e obrigatoriedade de serviços externos de auditoria auditarem... Não só o sistema em si, mas os processos envolventes, não é? Sim, certamente. O governo, na minha ótica se houver uma suspeita, tem que ter toda a liberdade para investigar. Ninguém está livre de proteger quando a suspeita é algo.... Criminal ou algo do género ou que vai impactar a segurança nacional, portanto como se diz.

Appendix B.13 – Answers provided for the Interview Question 13

	Interview Question 13: How is feedback from external stakeholders, such as regulators or the public, incorporated?
1	Eu acho que é sempre muito bem incorporado.Ou seja, eu acho que também poderia haver logo de forma automática algum resultado que saísse fora do esperado. Seria logo uma flagzinha para se averiguar mais tarde. Se não tipo uma caixa online de sugestões.
2	a opinião pública pode chegar porque tanto o cliente como a empresa está a desenvolver, acho que devem ter um departamento ou pelo menos algumas pessoas estejam encarregues em... De estar atentos isso e ver as redes sociais, às vezes pode-se chamar o departamento de culture outras vezes pronto enfim... depende sempre, mas existe esse departamento normalmente tanto de um lado como do outro. E acho que se calhar até faz sentido na fase de pronto do go to market haver essa awareness da equipa de que tem que estar atentos a esta parte também. Mas acho que também pode haver outras iniciativas. Por exemplo fazer entrevistas ou inquéritos, acho que pode haver esse tipo de coisas. E também auditorias que não sei se são esses reguladores. Mas pronto, não sei qual é a obrigatoriedade de auditorias neste caso. Mas pelo menos os auditores de IT acredito que estejam um bocadinho envolvidos. Eles também, se for uma auditoria externa acho que também fica público, penso eu, para os outros stakeholders externos poderem ver.
3	A meu ver, só na fase de testes, em que possas... há duas estratégias ou quando produzes, quando metes a coisa a correr, vais segmentando. Primeiro fazes para um grupo específico, tens um universo de mil e começas com 10, 20, e depois aquilo que correu bem para aqueles 10, 20 e metes mais 10, 20, e vais andando até, depois já estás com confiança passar quatro ou cinco vezes grande e depois já generalizas. Portanto, podes arranjar um grupo tipo em que faças esse envolvimento inicial. Há várias formas, podes fazer sondagens podes fazer inquéritos, seja por mensagem, seja por email, seja por o que for. Podes fazer público, podes criar uma, a partir de todo o produto, há de ter um site institucional ou empresa, ou o que for que esteja a ser utilizado. Pode ser sempre lá uma parte de sugestões e de avaliação ao produto. Depois podes ter também, imagina, fizeste um serviço, eu falo aqui neste caso, não sei se isto depois generaliza, mas fizeste um atendimento telefónico sobre um esclarecimento ou alguma coisa, podes naquele momento pedir que a pessoa faça a sua avaliação e tens o feedback.
4	Os independent evaluators, dependendo do trabalho que eles façam, ou seja, eu não sei se existe, mas imaginemos que existe uma ferramenta que é desenvolvida e tem de ser avaliado por alguma entidade externa fora empresa contratada e fora empresa owner da plataforma para validar se determinada regulamentação se encontra cumprida. Eu diria que este, se existir este.... Terá tanto ou mais responsabilidade qualquer um dos outros que mencionei anteriormente porque no fundo esta pessoa também funciona quase como um auditor. De resto, lá está este courts and regulators se for uma questão, por exemplo um aviso de abertura que é publicado e foi mal interpretado ou tem a responsabilidade imputada se for, por exemplo como acontece ao nível da da autoridade tributária e cabe-lhes um pedido vinculativo de esclarecimento e esse pedido é publicado e de forma não dúbia indica que se pode fazer determinada coisa e se há problemas então a culpa também passa a ser dessa entidade, mas depende da forma como a informação é reproduzida.
5	Sim, eu diria que muitas vezes são obrigados a incorporar, portanto, (...). Mas aí, pronto, estou a dizer advocates deviam ter alguma responsabilidade, dentro do possível porque se for algo que lhes é escondido não é responsabilidade tanto deles, se calhar. Mas. Se não tiverem algum sentido crítico e investigarem o produto que estão a publicitar também é responsabilidade deles. Ok data subjects, as pessoas a quem pertencem os dados diria que não estou a pensar em algum caso em que... quer dizer, agora até na Europa acho que como é que é no GDPR. Acho que é se tem o right to an explanation se for feito uma decisão já não sei como é que era o certo, tem direito de uma explicação sobre a decisão feita por um algoritmo que lhes afeta. Não sei se é assim tão simples há de haver outras constraints. Nesse caso a resposta era sim.
6	
7	
8	
9	Pois teria que haver uns mecanismos, não é? E, se calhar, isso vai acontecer para a frente. Para haver alguma limitação deste tipo de mecanismos que acabam por ter alguma influência na sociedade a certas pessoas dar acesso, neste momento, o que as pessoas têm. Já se viu que é um problema. E está-se a tornar perigoso.
10	O sistema de tickets (...). Externamente, acho que melhor bastante. A comunicação para fora da empresa é muito melhor se for feita através de tickets, e depois o ticket chega à empresa e alguém recebe alguém está responsável por receber esse ticket comunicar qualquer coisa nesse ticket ou entrega-lo a outra pessoa mas quando é mas isso lá está é quando quando o circuito de informação vem de fora para dentro.
11	

Appendix B.14 – Answers provided for the Interview Question 14

Interview Question 14: Do you think accountability for the AI program's outcomes changes over time?	
1	
2	Eu acredito que numa fase mais avançada do projeto não vai ser a equipa de desenvolvimento que vai estar atenta vai ser mais o cliente.
3	Eu diria que pode mudar, mas lá está. Porque é assim, quando se faz os requisitos tu podes passar o requisito de uma forma pouco clara ou que responda uma parte da necessidade, mas que haja ali outra parte que tens mais... Não tão direta, e depois aquilo pode ir com essas falhas, coisas pequenas, mas que às vezes pode depois não responder. Mas a responsabilidade vai alterando mediante da fase de evolução, sim, porque... há a responsabilidade inicial de cumprir com os requisitos do fornecedor, depois há uma fase de testes em que a responsabilidade acaba por ser mútua, em que tem que haver uma garantia de um lado que cumpriu os requisitos e uma garantia do outro em que avalia e diz sim ok, isto é aceitável porque está... perante os requisitos e perante a necessidade, está a corresponder. Aí é o misto e depois, a partir daí, passa para o lado de cá. A partir daí, é o cliente que é responsável e o que acontecer sobre o sistema, até pode ser por consequência de coisas que foram mal desenvolvidas, mas a partir do momento que passou para o lado de cá, a responsabilidade é nossa. Acabas por ter é que depois imputar isso, esses problemas à manutenção que tiveres depois atual. Eu acho que é uma evolução. Nesse caso, acho que acaba por ser um bocadinho assim também.
4	Se calhar altera só ligeiramente com o tempo porque uma fase inicial não está implementada diria que no fundo há sempre a presença da empresa a quem pertence a ferramenta ou seja, quem comprou o desenvolvimento ou quem adquiriu determinada ferramenta porque tem que ser responsável pelo tipo de dados que armazena. Normalmente quem desenvolve não armazena esse tipo de dados faz os mecanismos responsáveis pelo armazenamento dessa informação. A empresa subcontratada, eu diria que também nunca é isento a 100% de responsabilidade pelo que legalmente possa ser impunido. Se houver coisas quando, na legislação do país, por exemplo na UE sejam incutidos, eu diria que a empresa que desenvolve determinada ferramenta terá sempre imputado alguma culpa se essas regulamentações não forem cometidas. Independentemente se o projeto já esteja desenvolvido ou não. No fundo funciona como uma auditoria quase, a partir do momento em que eu atesto que aquela ferramenta se encontra de acordo com as regulamentações impostas para a área geográfica daquela empresa tenho que as cumprir e portanto terá essa responsabilização. Diria que era capaz de ser e o próprio utilizador final tem sempre, ainda que menos tem sempre alguma responsabilização também. A responsabilidade maior desde o inicio até ao fim do processo, diria que é sempre das equipas, mas no geral da empresa, que envolve na contratação, ou seja, a que não é contratada, que é o owner da plataforma.
5	Diria que não muda porque são responsáveis quando desenvolvem e têm de ser responsáveis se o problema tiver a ver com o desenvolvimento do modelo, quer passa muito tempo ou não. Se o modelo for biased ou a responsabilidade também que continua a ser deles, quem desenvolveu o modelo.
6	A responsabilidade é sempre da organização do executive management a organização que está fora está a prestar um serviço a responsabilidade. Isto é tomada de decisão, apoio a tomada de decisão portanto eu confio ou não confio naquilo que no sistema que eu estou a gerar na informação do sistema me está a dar para eu tomar decisões. Mas eu como executive management sou o único responsável pela decisão que tomo e estes é que são os grandes problemas do ponto de vista de responsabilidade deste tipo de sistema não posso dizer ah o modelo não treinou bem isto não acontece. A decisão que ele que toma e o resultado que o modelo nos dá não pode ser uma caixa preta ou seja é por aí aquela questão do explainable AI ou seja tem que ser capaz de justificar, não é só a resposta sim ou não. É como é que chegou a essa resposta, porque e quem é que é responsável por isso porque o cliente que está de fora pode dizer então mas porquê que me deram esta resposta? Porque e é isso a responsabilidade última é sempre da gestão do executive management.
7	You mean the whole team that, say, hire another company, another external company to rebuild, to reuse, to move, continue to monitor the, the products. Then by contract, uh, each team, he should be responsible for, for the product at the, at one moment. Once your contract with, uh, with the client has ended and he said, no, I don't work with you anymore. There's a process of you to help the new team to give the code, give the right access to the code to the new team. And once the transition is done, then you have no more or really less responsibility about that. Unless maybe if you made a really huge mistake, not by error or intentionally, then maybe the client will have to say a word to you. For example, if you left a backdoor in your code, or if you, by mistake, gave wrong data from algorithm. That makes your company lose a huge amount of money, then maybe you get sued by the client. But otherwise, once the headover has been done, the new team will take responsibility for the project now, from this point to the point of future.
8	Depende no início pronto no início, não é, quando acabamos por fazer a programação e mandar para lá o código existe um acompanhamento. Não é para verificar e ir verificando se realmente teve tudo bem se não houveram erros, pronto. Se não ficamos pronto não podemos estar a verificar constantemente todos os clientes que a gente tem. Então é quando existe problemas pois nós somos notificados e vamos lá e resolvemos isso. Creio que no nosso caso é quase sempre assim o cliente raramente se dá como culpado até mesmo quando utilizam... quando cai a responsabilidade deles não é. Que eles é que alteram as configurações que não deviam porque quando tens o software de casa aquilo é passível de alterar. É chegares lá não para visualização mas também para interação para dar os inputs e muitas vezes aquilo sofre alteração pelo próprio cliente e a responsabilidade cai na mesma sobre nós.
9	Normalmente é sempre quem desenvolve o software que fica responsável por corrigir esse tipo de coisas. A menos que seja, vá, um problema que não tenha nada a ver com o sistema, que seja uma coisa muito específica de instalação do cliente e se o cliente tiver condições para isso, que muitas vezes não tem. Porque não tem equipa especializada para tratar das situações ou para compreender qual é o problema. Mas o que acontece quase 100% das vezes, mesmo sendo coisas que o cliente pode resolver nas próprias instalações com o staff que tem, remete-se sempre para este lado primeiro. Há uma garantia. Muitas das vezes as garantias até são negociadas com o próprio cliente dentro do contrato. Há contratos em que o cliente tem um contrato mais efetivo, pagou logo à cabeça para ter um pacote ali pronto para ter uma assistência mais rápida. Mais celebre, com um conjunto de serviços mais abrangentes e depois há outros clientes que não querem isso, não é? Que preferem navegar à vista. Portanto depende muito do cliente. O mínimo é sempre as regras nacionais que nós temos de relação às garantias dos softwares. No nosso caso, nós até damos um bocadinho mais porque a gente trabalha com sistemas que se pararem o cliente perde a produtividade da empresa. E então damos sempre um bocadinho, damos sempre um pouco mais. Em termos do software, a gente

	quase que dá uma garantia vitalícia. Pronto, até o próprio software se extinguir, não é? Porque a gente o descontinuamos. Normalmente a gente está lá sempre prontos para resolver alguma coisa.
10	
11	Ou seja, as empresas celebram estes contratos de licenciamento para monitorização e manutenção, mesmo por causa disso, e também para suporte à aplicação e... Outro tipo de eventuais novas features que queiram... alguns bugs que possam ter encontrado que não encontraram durante a fase de testes imagino que isto é um produto que vende a um cliente e esse cliente revende a outro, portanto isto é sempre quem quer que a empresa desenvolva é quem tira a manutenção. Eu acho que se mantém, porque é um compromisso. Se é um compromisso que a empresa estabelece com o cliente não tem que ter mais ou menos. Portanto pode, se calhar a nível absoluto diminuir porque o número de pessoas responsáveis diminui porque é um produto já final, mas vai haver sempre pelo menos um ou mais responsáveis que garantem a manutenção do produto ou serviço. Acho que a esse nível, mais macro, continua igual, acho que é constante. A não ser que a empresa acabe por comprar a solução e não queira mais e isso desaparece. Mas creio que seja constante, creio que a responsabilidade não diminui.

Appendix B.15 – Answers provided for the Interview Question 15

	Interview Question 15: If you had to summarize everything into 2 or 3 main stakeholders, who would they be?
1	Então será o platform owner, o product manager, no terceiro eu diria a equipa toda aí em baixo.
2	Por exemplo, o end-user tem muito interesse, mas não tem responsabilidade. A responsabilidade acho que é das duas equipas de data, tanto do lado da equipa de desenvolvimento e do cliente e das equipas de... ou melhor das direções executivas tanto do lado como do outro acho que são os principais responsáveis. Mas eu diria que acho que é a equipa do cliente, porque apesar de tudo, se eles não quisessem que isto avançasse para a frente, eles podiam dizer que não, mesmo que tivesse sido mal construído. Por isso acho que no final a maior responsabilidade mesmo é a equipa do lado do cliente. O data controller e o model maintainer. Quem vai dar a cara no fundo é o Project Owner. Eu digo isto plus as equipas executivas.
3	Portanto eu diria em paralelo ao gestor de projetos, tens de ter aqui de um lado e outro. Portanto, da parte dos testes há de ser lá a pessoa responsável pelos testes (...). Podes falar depois noutros aspetos, tens depois a mais superior, tens a parte da administração que também tem um peso grande aí, porque não é fácil gerir quando tens, por exemplo, pressões.
4	Se considerarmos a constituição como aqui está, os principais eu diria que são o platform owner do lado do cliente, o product owner do lado da empresa contratada e depois provavelmente o executive manager da empresa owner também.
5	Pode ser um bocado biased mas se calhar ia dizer data scientist ali daquela linha, depois tipo tipo project funder, executive management também e depois aqui ... imagina o impacto depois é medido nos end users mas se calhar em termos de importância de desenvolvimento e de stakeholder não está muito neles. Portanto se calhar o terceiro pode ser ... o courts and regulators (...). Pode ser, pode ser acho que afeta um bocado tipo data scientists, project fund, executive management, court and regulators, pode ser.
6	com o máximo de responsabilidade e com o máximo de de tudo que é o executive manager da organização que vai usar o sistema. Porque é ele que é ultimamente responsável por este sistema e pelas decisões que depois são tomadas com base na informação que o sistema me dá. Não vale a pena avançar portanto e isso é o que nós sabemos já é um fator crítico de sucesso que sabemos de todos os sistemas de apoio à decisão. Todo o desenvolvimento, ou seja, se nós podemos ter uma equipa do lado aqui do outro lado da equipa da organização que desenvolve. Ok temos que ter utilizadores que depois saibam usar temos que ter uma equipa. Ou seja, há muita coisa que tem que funcionar. quando estamos a falar de fatores críticos de sucesso deste tipo de sistemas há muita coisa. Os dados têm que ser bons, temos que cumprir as questões legais e éticas porque senão ficamos com uma enorme dificuldade do ponto de vista de incumprimento...
7	Usually, on an IT project, stakeholders are the users, the clients, because it can be different. The team who produces the products. That's who the one I work with. And now you have extra stakeholders related legal issues so basically yeah. Client, user.
8	as pessoas principais o programador, creio que a pessoa responsável também, que é a pessoa que vai dar a aprovação, e normalmente pronto nós acabamos por ter muito dessas, mas existe depois a pessoa acima do meu responsável, neste caso que é que acaba por ter o contacto com o cliente e aprovar realmente não é, o que é que o cliente quer. Ver se está tudo em conformidade com aquilo que o cliente pretende ou não. É um pouco assim que acontece.
9	É o cliente. Se é o teu projeto. Se calhar mais o gestor do projeto do que o comercial. E depois é a equipa de desenvolvimento. Tendo o cliente a maior responsabilidade. O gestor tem uma responsabilidade compartilhada com o cliente. E a equipa de desenvolvimento tem menos responsabilidade.
10	Os absolutamente essenciais o CEO é absolutamente essencial, sem ele nada avança não há uma tomada de decisão portanto ele é absolutamente essencial. Depois a pessoa que idealiza o produto é a segunda pessoa que é essencial porque se ele não idealizar nada não se faz e não se cria. E depois há aqui umas pessoas que podemos dizer que são essenciais... elas na verdade não são essenciais porque outras podem assumir as funções dela mas... a seguir vem o corpo técnico pronto digamos assim a parte a parte dos data analysts dos data scientists da programação porque sem eles o programa não é executado. A parte do project manager e do product owner a pessoa que idealiza que a pessoa que idealiza o produto cumpre bem porque se for necessário cumpre bem esse tipo de de função e então pode saltar esses dois.
11	stakeholders principais, sem dúvida o gestor de projeto e a empresa talvez esteja a contratar o serviço não? Sim. Talvez um platform owner ou... exato.

Appendix B.16 – Answers provided for the Interview Question 16

Interview Question 16: What would you suggest to improve the involvement of these 2 or 3 key stakeholders?	
1	O envolvimento poderia ser melhorado através de uma definição melhor dos requisitos logo a priori. Pronto depois eu diria só então talvez uma equipa extra só para fazer testes para garantir que está tudo bem, uma equipa de manutenção.
2	os executivos têm que ir fazendo meetings ou pontos de situação regulares com os project owners e com os managers do project eventualmente. Têm que garantir também que a equipa que trata de avaliar a opinião pública, etc, e dos end users, tem que garantir que eles estão atentos. reunirem de forma organizada e saberem onde é que está a informação por exemplo se a equipa reunir diariamente com o project manager, o project manager está por dentro, se o project manager reunir diariamente com o project owner e com o data controller, por exemplo eles também vão estar por dentro e quando eles forem ter a discussão, o ponto de situação com a executiva e vão estar alinhados. Eu acho que no fundo é isso, é manter a organização e, isto mais se calhar para as equipas funcionais do lado do cliente e para a equipa que está a desenvolver, tem que manter tracking sempre das coisas. Daí lá está, é importante a plataforma que escolhes para organizar tudo. Seja ele o Trello, ou seja, a plataforma que tu quiseres e-mail ou não, eu acho que tem que estar tudo muito organizado de forma a haver tracking de tudo e que as coisas mais importantes passem para a executiva.
3	Para mim isso é claro, tem que haver vários pontos de situação, tem que haver vários... A comunicação é sempre acima de tudo e muitas vezes é o que falha nas organizações. Porque é sistemas hierárquicos depois é difícil o acesso e tens que ir lá e pôr-se o senhor doutor, por favor, tal, tal, engenheiro isto ou aquilo. E depois há um tema qualquer e para chegar lá já vai... vai-se cortando no que se vai dizendo, ali contas um bocadinho, quando chega lá acima já é só uma coisinha de nada e pronto. Portanto tem que haver a transparência suficiente e tem que haver a possibilidade de haver esta transparência e pontos de situação capazes de ir alinhando a estratégia para chegar ao final, que acho que a grande dificuldade é essa.
4	Normalmente já existe comunicação aberta quando é para o desenvolvimento de uma coisa do género.
5	
6	
7	Between the clients and the teams there should be in regular contacts to understand the needs to choose to stay in touch in time because, because usually they have needs they are changing. (...) because our identity has changed, our logo has changed. So there should be regular contacts between the clients and the team of the uh technical team. The legal team, (...) they should also also either supervise all project rising a country for example and make sure they create law for that or rising a company from either the client side of the technical side that should just make sure that the demand are ethical and not breaking the laws. Maybe there are two layers for that for that? One internally and one external? One making sure that they're not breaking the law and the other one creating or getting the law to make sure that everything is everyone is safe.
8	
9	Que houvesse uma documentação prévia, do que é que ia ser tudo feito. E a equipa de desenvolvimento tem outro tipo de sensibilidade para certas situações. Ter uma documentação detalhada do que é que era a ser feito. E conseguir identificar problemas antes deles acontecerem. Seria a única maneira de isso acontecer. Outra coisa é haver um terceiro elemento. Que neste caso, não sei se temos alguém em Portugal com capacidade para isso. Seria alguém que dominasse a parte legal. Um advogado, por exemplo. Que olhasse para um projeto desse tipo e conseguisse olhar, oh, estes dados podem ou não ser usados. Quais são as consequências de se usar este tipo de inteligência artificial.
10	Pois a parte da comunicação entre as pessoas é uma situação muito complicada porque o que é óbvio para um não é óbvio para o outro e depois cria-se aqui o tal gap. Para um dizer aquilo é completamente desnecessário para outra pessoa não ter ouvido aquilo, ficou perdida. Portanto aí é é tentar-se criar uma estrutura de comunicação uma obrigatoriedade daquilo que se vai comunicar e de quando é que se vai comunicar. É que é não deixar essa parte ao bom senso. É criar um método este é um método de comunicação da nossa empresa, estes tópicos não são não se podem omitir nunca, portanto fala-se sempre disso não interessa quantas vezes já tínhamos ouvido fala-se. O sistema de tickets internamente, não o acho que melhor, sinceramente. Quando é é por exemplo hierarquicamente dentro da empresa funciona bem. Quando a informação vai descendo vai do superior vai da hierarquia superior e vai para a hierarquia inferior. Agora quando a informação tem que subir eu acho que é muito mais difícil.
11	Se o project manager ou o project owner são os stakeholders que devem ter maior conhecimento de sempre sobre a solução que querem, portanto eles estão no topo da responsabilidade, não é? Eles conhecem o projeto de trás para a frente e se não conhecem devem não conhecer. E depois é a empresa que contrata o serviço, portanto o tal platform que falou. Portanto se empresa contrata o produto e serviço, quer exatamente e sabe exatamente o que quer, também deve ser high stakes de responsabilidade, não é?

Appendix B.17 – Answers provided for the Interview Question 17

Interview Question 17: Would you like to share any limits, challenges or interesting topics about AI systems?	
1	
2	Quero acrescentar que eu acho que quem vai definir realmente as responsabilidades é a legislação e a opinião pública. Agora, eu não tenho noção se já existe. Claro que deve existir não é? Legalmente deve existir forma de dizer que realmente quem é que é culpado e quem é que é mais culpado e pronto. Mas eu acho que isso não está assim tão no dia-a-dia da equipa. Acho que no dia-a-dia da equipa, a equipa sente-se sempre responsável. Por tudo.
3	Imagina, isto já aconteceu para trás, agora acho que já não se usa tanto, mas era aquela ótica de isto está planeado para março mas tem que estar feito no fim do ano. E depois, em vez de se desenvolver uma coisa que funcione bem e que tenha lá as coisas todas, aquilo fica tudo enrolado, entregas na mesma, mas depois aquilo está sempre a dar bugs por todo o lado e rebentar. Portanto acaba por também ter um papel relevante na medida em que tem que dar. Tem que gerir a parte lá superior da administração e afim, mas tem que garantir que se consiga aqui um consenso e timings adequados ao desenvolvimento do projeto. Portanto há aqui várias fases importantes.
4	
5	
6	Exatamente pronto e aí que tem que ser muito bem explicado do ponto de vista da gestão porque às vezes porque às vezes isto é importante o seu trabalho é importante para explicar aos gestores que isto não é um computador que se compra é um computador é muito mais complexo que isto e as implicações as implicações que tem do ponto de vista legal e do ponto de vista de responsabilidade é muito grande.
6	Sem se as pessoas aperceberem que estamos a ir num momento em que em que as coisas são verdadeiramente críticas porque reparem antigamente nós usávamos inteligência artificial e toda esta parte da automatização é robótica noutras coisas destas por exemplo eu vou pôr um robô numa cirurgia a fazer a substituir o cirurgião este sistema não foi bem validado voltando atrás não foi bem testado e validado qual é a taxa de erro que é aceitável? Mas depois numa aplicação de crédito, por exemplo, já não há problema. Ok, então, mas vamos lá ver. Não pode ser assim e crédito ainda é uma coisa que não tira a vida. Agora cada vez mais isto está a ir para situações mais abrangentes e mais complexas. E nós temos de perceber, uma das coisas que a gente fala da validação deste tipo de modelos é e o 1% que falhou? Qual é o impacto desse 1% ? Estamos confortáveis com o impacto que isso causa? No limite a responsabilidade de tudo é da liderança de tudo. Ownership ownership e accountability é de tudo.
7	One thing I heard from my friend who are in the tech companies is that the challenges of the future will be trying to repair everything that AI has done. That means that we won't be in shortage of work as developer we just will need to have more experienced people to person to work on the project. (...) and what we saw a lot of people, a lot of people are anticipating is that in the future, you may have to repair and repair everything like that. Have a review of the code to repair the repair. It's like giving you, like you try to build a puzzle and you're missing one one piece but you take it the puzzle from another piece for another puzzle that fits in your piece, it fits, it's complete but yeah it's not so you're not the right one but it fits.(...) let's move on but you have to review you, if we have to review what's it gave you it's like when you're writing an email we can ask him for example write an email to a company… here we can give you an email but you have to review it to or talk to spend time with it to ask him to update it to personalize and at the end you have to redo it in your own words to make it less you know generic and more targeted to the to the audience.
8	Acontece ok acontece e pronto é como tudo na vida não é. Quanto mais tempo tens para entregar alguma coisa melhor fica. Pois quando não te dão muito tempo nem muitas hipóteses e querem as coisas muito em cima da hora, é propício ao maior número de erros e consequente teres que lidar com eles mais tarde.
8	Eu utilizo com muita frequência, mas compreendo todos os perigos associados. Eu ia dizer, especialmente para a malta (...) muito facilmente enganada. E isso sim, claro, representa um perigo. E há N imagens na internet, há N publicações na internet que sim, se tu tiveres a ler um artigo, tu acreditas, porque é plausível, não é? Sim, nesse caso torna-se complicado gerir. E depois também as bases de dados que está a usar carecem de updates. Vá, aquilo usa, o que é? Se calhar, vá. O motor do chat GPT, o três e meio, usava a informação da que há três anos atrás, algo assim, creio, se não estou em erro. Por isso, pronto, não estão atualizados, não é? E por isso há muitas respostas que acaba por dar, que também estão erradas.
9	Se a gente tiver um, normalmente esse tipo de motores autoalimentam-se, ou se têm a memória. E precisam ter a memória para se recordarem das conversas e seguirem o raciocínio, não é? Outra coisa que às vezes pode acontecer é a pessoa começar a introduzir dados pessoais. Fazer uma pergunta simples, introduzindo a morada dele, o nome. Isso vai ficar lá. Depois, o que é que a OpenAI faz com aquilo? Nós já não sabemos. Neste momento há várias, já deves saber disso, há várias coisas relacionadas com a proteção dos dados. E possivelmente há situações em que... O chat GPT vai ter que andar para trás. Porque neste momento está a utilizar a informação de várias fontes que nunca deram permissão para isso. E esse é um problema porque as pessoas desconheciam também esse facto, não é? As pessoas acharam que aquilo era uma coisa tradicional, não tinham noção do que é que se passava. E foram introduzindo dados. Eu não te sei dizer qual eram as regras de utilização, o disclaimer que eles lá tinham na altura. Mas uma coisa que eu te posso garantir, ninguém o lê. E temos o lama que é o do Facebook. Claramente foi treinado com as coisas que as pessoas deixaram no Facebook. As coisas que estão no Facebook por norma são pessoais. E no entanto, nada evitou que uma entidade como o Facebook utilizasse esses dados para fazer o treino do modelo.
9	Estamos a falar, estamos a falar de um negócio. Um negócio em que o tempo é o dinheiro. Porque este tipo de desenvolvimento de software não gastas, tirando energia elétrica. O maior peso é o tempo, então os projetos têm a tendência a ser mais complexos, mais curtos possíveis para entregar o produto o mais rápido possível. E isso depois vai sempre causar alguns constrangimentos. Mas lembrar-te-ás daquilo que eu te disse, na altura da qualidade, de fazer verificação, dos testes e da qualidade, é por isso que o cliente depois acaba por ser envolvido. Às vezes poderia não querer ser. Mas tu te lembrares, vais-te lembrar que o telemóvel tem atualizações constantes. Os computadores têm atualizações constantes. Agora temos os carros com atualizações. E isso vem tudo deste facto. As coisas não são desenvolvidas até a exaustão, não são testadas até a exaustão, antes de irem

	para o mercado. A própria tecnologia permitiu que isso acontecesse. As coisas colocadas no lado do cliente e depois isso corrigir os problemas conforme eles aparecem. Portanto, é uma realidade do software, infelizmente, que cada vez é mais.
9	Mas qualquer pessoa que tenha trabalhado um pouco com a inteligência artificial sabe que eles estão muito longe de serem perfeitos. Têm até uma taxa de sucesso de 80%. E o resto é falhas. Falhas às vezes que são críticas e que têm uma diferença de valores muito distante uns dos outros. Porquê? Ou porque não recebeu a informação toda que devia ter, não foi alimentado corretamente, ou mesmo porque a tecnologia não está o suficientemente maturada para fazer isso. Mas sim, vamos ter esse problema, vamos começar a ter muita informação, pessoas a receber o output desses sensores todos e desses dispositivos e vamos ter que começar a lidar com esse tipo de coisas. Que no caso da saúde mete-se muito a responsabilidade de quem faz a recolha dos dados, que é uma coisa pessoal, de quem vai guardar estes dados. Há uma coisa que tu aqui não falaste, mas já pode haver um terceiro elemento que tu não tens aqui, pode ser a AWS, pode ser a Google, pode ser a Microsoft, que têm uns servidores e que vão ter essa informação toda guardada da aplicação e eles acabam por ser os detentores dos dados e muitas vezes eles fazem cópias de segurança destes dados e nem o cliente neste caso, nem o desenvolvedor tem a responsabilidade daquilo e pode haver alguma fuga de dados, uma coisa qualquer, e estes dados irem parar a um sítio onde não devem. A parte da base de dados em si, o usuário destes dados também acaba por ter alguma responsabilidade.
10	(Legal and Ethical) É assim existirá essa pessoa existirá, mas existirá em empresas extremamente grandes, existirá numa google existirá numa microsoft e existirá em empresas desse género agora se calhar provavelmente nem no nem no unicórnio português da atualidade existirá uma função similar.
10	Eu confesso eu confesso que não o uso muito mas daquilo que usei e aquilo que usei aqui em casa obviamente que detetei lacunas basta a pessoa virar um pouco a cara às vezes a inteligência artificial já não a reconhece e é preciso tentar de alguma maneira colmatá-la a situação é assim do género. Nós sabemos que que a visão computacional e o reconhecimento de objetos e de feições de pessoas é extremamente complexo. Continua a ser hoje um desafio enorme pôr as máquinas a ver, mas sim detetei essas lacunas basta às vezes haver uma mudança subtil e a inteligência artificial já não tem a competência para o fazer enquanto que um ser humano qualquer em dois segundos reconheceu a cara que ali estava.
11	futuro sem dúvida a generalizada, o AI. E depois é um bocado irónico porque é a mais esperada e a mais temida, porque já estamos a atingir níveis de self-awareness quase um bocado assustadores. No entanto, novamente aqui apelando à responsabilidade, quem desenha sistemas e quem os concebe sabe perfeitamente o que está a fazer estas máquinas não vão ganhar vida por si próprias a não ser que nós queiramos que ela se reinventa ela própria e para isso ela tem que ter alguma instrução para tal. Isto é tudo puramente estatístico e puramente factual, que não é uma black box, apesar de dizem que é o pai da AI dizer que eles ganham consciência e ganham nada. Não, é verdade. Quem desenha estes sistemas sabe exatamente o que está a fazer, portanto, se eles eventualmente chegarmos ao nível do AI foi porque alguém assim o quis, alguém bastante poderoso. Portanto, esse talvez seja um bocadinho de balanceamento entre aquilo que é esperado, aquilo que vai ser o futuro e aquilo que talvez possa ser temido. Não tanto de uma crise global, mas no que diz respeito ao ataque aos empregos, à diminuição brutal da carga de trabalho das grandes empresas e obrigando a muitos indivíduos a requalificarem-se para se manterem ativos na sua vida profissional. Eu penso que isso seja a ameaça barra oportunidade, pois quem está nestas áreas vai ver ainda mais oportunidades, portanto isto tem um pau de dois bicos, não existe, mas penso que seja isso assim ultimamente.
11	Não as pessoas sabem. As pessoas sabem o que é que estão a fazer. A informação é que não passa cá para fora, está a perceber? É o pequeno núcleo de quem desenvolve. Outra coisa é as pessoas que estão cá fora a ver o que é que está desenvolvido. Portanto não é a fraca noção da equipa desenvolve, é a fraca noção de quem está cá fora e não percebe destes temas, ou pelo menos não se inteira do que é que funciona tecnicamente por trás. Isso é que pode ser um risco o risco leva... O desconhecimento leva ao mau julgamento, o mau julgamento leva a atitudes que podem ser nocivas. Portanto, a falta de noção não é de quem desenvolve, é de quem está de fora a ver. As pessoas que trabalham a área sabem exatamente o que está a acontecer e muitas das informações que saem dessas pessoas cá para fora já vêm manipuladas por isso é que aquele prémio Nobel da física e da Jeffrey e qualquer coisa... sim, esse senhor eu acredito que aquela entrevista tenha sido se calhar a pior entrevista que alguma vez deu não, ele sabe exatamente o que é está por trás. Portanto, aquilo foi mais para dizer às grandes empresas que estejam quietas do que propriamente assustar as pessoas, mas ele conseguiu isso num discurso misto porque ele sabe exatamente, ele diz que é uma caixa negra, mas não é, portanto ele sabe exatamente o que é que está a acontecer. Não há magia, não é nada mágico, ninguém ganha consciência é tudo fictício, esse tipo de conversa, mas a fraca noção, a fraca noção é mesmo quem assiste fora e não para quem está lá dentro para quem está lá dentro sabem exatamente o que é que está a acontecer.

Appendix C – Data Protection Statement in Portuguese - Informed Consent

O presente estudo surge no âmbito de um projeto de investigação a decorrer no ISCTE – Instituto Universitário de Lisboa. O estudo tem por objetivo explorar o conceito de responsabilidade nos stakeholders envolvidos em programas de Inteligência Artificial dedicados à negociação.

O estudo é realizado por Leonor Patrício Mendes, com o endereço de email leonorpatriciomendes@gmail.com, que poderá contactar caso pretenda esclarecer uma dúvida ou partilhar algum comentário.

A sua participação no estudo, que será muito valorizada pois irá contribuir para o avanço do conhecimento neste domínio da ciência, consiste em responder a um conjunto de perguntas para que se possa interpretar um diagrama realizado pela investigadora, num período de 15 a 20 minutos. Não existem riscos significativos expectáveis associados à participação no estudo.

A participação no estudo é estritamente voluntária: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação em qualquer momento sem ter de prestar qualquer justificação. Para além de voluntária, a participação é também anónima e confidencial. Os dados obtidos destinam-se apenas a tratamento estatístico e nenhuma resposta será analisada ou reportada individualmente. Em nenhum momento do estudo precisa de se identificar.

Declaro ter compreendido os objetivos de quanto me foi proposto e explicado pela investigadora, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora, pelo que aceito nele participar.

_____ (local), _____/____/_____ (data)

Nome:

Assinatura: